

# Mathematical Formulas and Other Valuable Knowledge

that I have found Useful for Myself  
and Decided to Write Down

Ott Toomet

Seattle, October 30, 2019

This is a collection of formulas, definitions, and other math stuff I have had to look up, find out, struggle with, and decided to write down for the future reference. The pdf version is available at [Github Pages](#) and the (mostly latex) source [on Github](#).

Feel free to use and copy the formulas here. But be careful – there is probably a number of errors and misunderstandings!

The cheatsheet started around 1999 when I studied economics at Tartu University and wrote a sheet (yes, a single sheet ;-)) of formulas to help my coursemates with math. There are still many pieces of text from that time, in particular those related to limits and Taylor's approximation. The pages filled with Tobit and duration models' Hessians are leftovers from my Master and PhD thesis. I wrote it in Estonian back then, and I don't intend to translate it just for translation's sake. Only if I am rewriting an old piece of text I'll do it in English.

Have fun!

Ott

Seattle, October 30, 2019

## Contents

<b>1</b>	<b>Geometry</b>	<b>5</b>
1.1	Koordinaatteisendus	5
1.1.1	Pinnaelement koordinaatteisendusel	5
1.2	Kolmnurk	5
1.2.1	Seos kolmnurga külgede ja nurkade vahel	5
1.2.2	Kolmnurga pindala	5
1.3	Trigonometrics	5
1.4	Joone kõverus	5
1.5	Supporting Hyperplane	6

<b>2</b>	<b>Functions</b>	<b>7</b>
2.1	Elementary functions	7
2.2	Algebraic Functions	7
2.3	Other functions	9
<b>3</b>	<b>Calculus</b>	<b>10</b>
3.1	Logarithm	10
3.2	Limits	10
3.2.1	Limits in general	10
3.2.2	Other limits	10
3.2.3	Limits Involving Factorial	11
3.3	Boundedness and Convergence	11
3.4	Differentiation	13
3.4.1	Simple Derivatives	13
3.4.2	Directional Derivative	13
3.4.3	Subderivative	13
3.4.4	Jacobian Matrix	13
3.4.5	Ümbrikuteoreem ( <i>envelope theorem</i> )	13
3.4.6	Normal Density Related Derivatives	14
3.4.7	Derivatives of Gamma Function	14
3.4.8	Differentiation of Sums	14
3.4.9	General Differentiation Rules	14
3.5	Integration	15
3.5.1	Simple Integrals	15
3.5.2	Integrals related to probability distributions	15
3.5.3	Other integrals	17
3.5.4	Üldised integreerimise reeglid	17
3.5.5	Laplace'i teisendus	18
3.5.6	Numbriline integreerimine	18
3.6	Differential Equations	19
3.7	Optimization	20
3.7.1	Second-order conditions	20
3.7.2	Optimal control (dynamic optimisation)	21
3.7.3	Newton-Raphson maximization	21
<b>4</b>	<b>Algebra</b>	<b>23</b>
4.1	Mõisted	23
4.2	Sets	23
4.2.1	Definitions	23
4.2.2	Basic Operations	23
4.3	Simple Algebra	24
4.4	Taylori rida	24
4.5	Summade astmed	26
4.6	Linear Algebra	27
4.6.1	Matrix	27
4.6.2	Determinant	27
4.6.3	Inverse Matrix	27
4.6.4	Matrix norm	28
4.7	Differentiation of Matrices and Vectors	28
4.8	Võrratused	30

4.8.1	Hölder's inequality	30
4.8.2	Jensen Inequality	30
4.8.3	Triangle Inequality	30
4.8.4	Cauchy-Schwartzi võrratus	30
4.8.5	Inequalities, containing exponent	30
4.9	Differential Equations	31
4.9.1	Linear Equations with Constant Coefficients	31
<b>5</b>	<b>Probability and Statistics</b>	<b>32</b>
5.1	Combinatorics	32
5.2	Random Variables	32
5.2.1	General Concepts	32
5.2.2	Sufficient statistics	32
5.3	Stochastic Boundedness and Convergence in Probability	32
5.4	Distributions: General Concepts	33
5.4.1	Sõltumatud juhuslikud muutujad	33
5.4.2	Expectations	33
5.4.3	Central and Non-Central Moments	33
5.4.4	Characteristic Function	34
5.4.5	Momendifunktsioon ( <i>moment generating function</i> )	34
5.4.6	Kumulandifunktsioon ( <i>cumulant-generating function</i> )	34
5.4.7	Probability Generating Function	35
5.4.8	Distribution Function of a Function of Random Variable	35
5.4.9	Information and Entropy	35
5.4.10	Kullback-Leibler divergence	35
5.4.11	Mutual Information	35
5.5	One-dimensional discrete distributions	36
5.5.1	Bernoulli	36
5.5.2	Binoomjaotus	36
5.5.3	Diskreetne jaotus	37
5.5.4	Geomeetiline jaotus ( $Geo(p)$ )	37
5.5.5	Multinoomjaotus	37
5.5.6	Poissoni jaotus	37
5.5.7	Skellam Distribution $PD(\lambda, \delta)$	38
5.6	1D Continuous Distributions	39
5.6.1	Beta distribution $\mathcal{B}(a, b)$	39
5.6.2	EkspONENTjaotus $\mathcal{E}(\theta)$	39
5.6.3	F-jaotus $F(n_1, n_2)$	40
5.6.4	Gammajaotus $\mathcal{G}(\alpha, \beta)$	40
5.6.5	Hii-ruut jaotus $\chi^2(k)$	42
5.6.6	Inverse gamma distribution	42
5.6.7	Laplace distribution	42
5.6.8	Log-normal $LN(\mu, \sigma^2)$	42
5.6.9	Log-ühtlane jaotus	43
5.6.10	Logistic Distribution	43
5.6.11	Normal Distribution $N(\mu, \sigma^2)$	43
5.6.12	Pareto Distribution	45
5.6.13	Pööratud normaaljaotus	45
5.6.14	t-Distribution	45
5.6.15	Triangular Distribution	46

5.6.16	Type-1 Extreme Value Distribution $EV_1$	46
5.6.17	Type-2 Extreme Value Distribution $EV_2$	46
5.6.18	Uniform Distribution $Unif(a, b)$	46
5.6.19	Weibulli jaotus	47
5.7	Multivariate Continuous Distributions	48
5.7.1	Dirichlet distribution	48
5.7.2	Multivariate Normal $N(\mu, \Sigma)$	48
5.7.3	Wishart distribution	50
5.8	Distribution Families	51
5.8.1	Exponential Family	51
5.8.2	Stabiilne pere	51
<b>6</b>	<b>Estimators</b>	<b>52</b>
6.1	M-Estimators	52
6.1.1	Variance	52
6.2	Maximum likelihood	52
6.2.1	Definition	52
6.2.2	Information matrix	53
6.2.3	Relationship to Kullback-Leibler divergence	53
6.3	Generalized Method of Moments	53
6.3.1	Optimal Weighting Matrices and Asymptotic Variances	54
6.4	Entropy Distance	54
6.4.1	Entropy Distance	54
<b>7</b>	<b>Stochastic Processes</b>	<b>55</b>
7.1	Autoregressive (AR) Processes	55
7.2	Hulkumine	55
7.2.1	Hulkumine vastu barjääri	55
<b>8</b>	<b>Statistilised mudelid</b>	<b>56</b>
8.1	Tobit-2 model	56
8.2	Tobit-2 Model with Binary Outcome	58
8.3	Tobit-5 Model	60
8.3.1	Heckmani kahesammuline hinnang	60
8.3.2	Maksimum-laiklikhuud hinnang	60
8.4	Kestusmudelid	65
8.4.1	Kaplan-Meieri hinnang	65
8.4.2	Multiplikatiivne mittevaadeldav heterogeensus	65
8.4.3	Tükati konstantne hasart	66
8.4.4	Intervallandmed	69
8.4.5	Parametriseerimine	73
<b>9</b>	<b>Algorithms</b>	<b>76</b>
9.1	Arrays	76

# 1 Geometry

## 1.1 Koordinaatteisendused

### 1.1.1 Pinnaelement koordinaatteisendusel

Üleminekul koordinaatsüsteemilt  $(x, y)$  uutele koordinaatidele  $(u = u(x, y), v = v(x, y))$  teisenevad väikesed koordinaatide sihilised lõigud nii:

$$\begin{pmatrix} du \\ dv \end{pmatrix} = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix}, \quad (1.1.1)$$

kus paremal poolel esimene on vastavate osatuletiste maatriks (jakobiaan  $\mathcal{J}$ ). Pinnaelement avaldub

$$dS = dx dy = \frac{du dv}{|\det \mathcal{J}|}, \quad (1.1.2)$$

kus  $|\det \mathcal{J}| = du dv \sin(\widehat{u, v})$  on kahe lühikese lõigu  $du$  ja  $dv$  poolt defineeritud elementaarpindala. Analoogiline seos kehtib ka kõrgemate mõõtmete puhul.

## 1.2 Kolmnurk

### 1.2.1 Seos kolmnurga külgede ja nurkade vahel

$$a^2 + b^2 - 2ab \cos \gamma = c^2. \quad (1.2.1)$$

Täisnurksel kolmnurgal, kui  $\gamma = 90^\circ$ , kehtib *Pythagorase teoreem*:

$$a^2 + b^2 = c^2. \quad (1.2.2)$$

### 1.2.2 Kolmnurga pindala

Kui kolmnurk on tasandil antud kolme punktiga  $(0, 0)$ ,  $(X_A, Y_A)$  ning  $(X_B, Y_B)$  siis kolmnurga pindala on

$$S = \frac{1}{2} |X_A \cdot Y_B - Y_A \cdot X_B| = \frac{1}{2} \text{abs} \begin{vmatrix} X_A & Y_A \\ X_B & Y_B \end{vmatrix}. \quad (1.2.3)$$

Tõestus: joonista kolmnurk välja ja vaata, millised pindalad kirjeldab determinant.

## 1.3 Trigonometrics

### Trigonometric Identities

$$\cos(\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta$$

## 1.4 Joone kõverus

Parameetriselt antud joone kõverus:

$$k = \frac{x' y'' - y' x''}{(x'^2 + y'^2)^{3/2}} \quad (1.4.1)$$

## 1.5 Supporting Hyperplane

*Supporting Hyperplane* of a set  $S$  in Euclidean space  $\mathbb{R}^n$  is a hyperplane that satisfies:

- $S$  is entirely contained in one of the two closed half-spaces bounded by the hyperplane;
- $S$  has at least one boundary-point on the hyperplane.

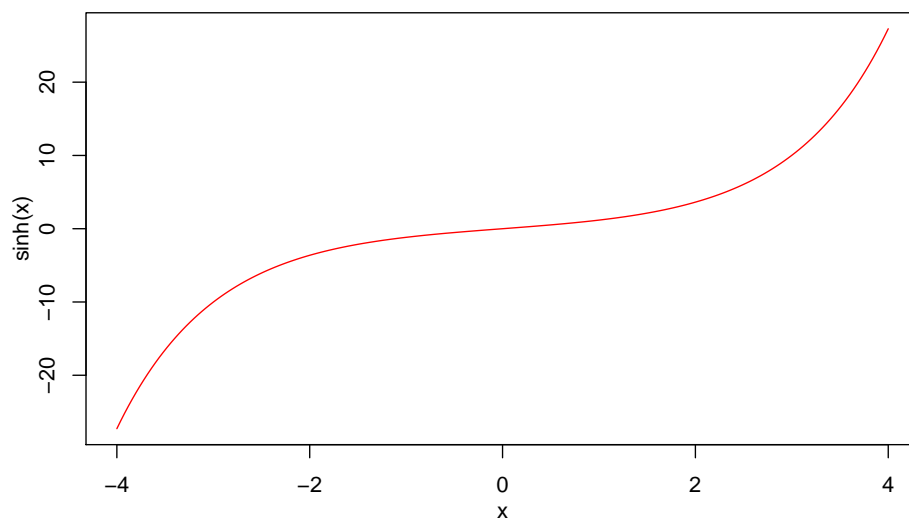
Supporting hyperplane may not exist if  $S$  is not convex.

## 2 Functions

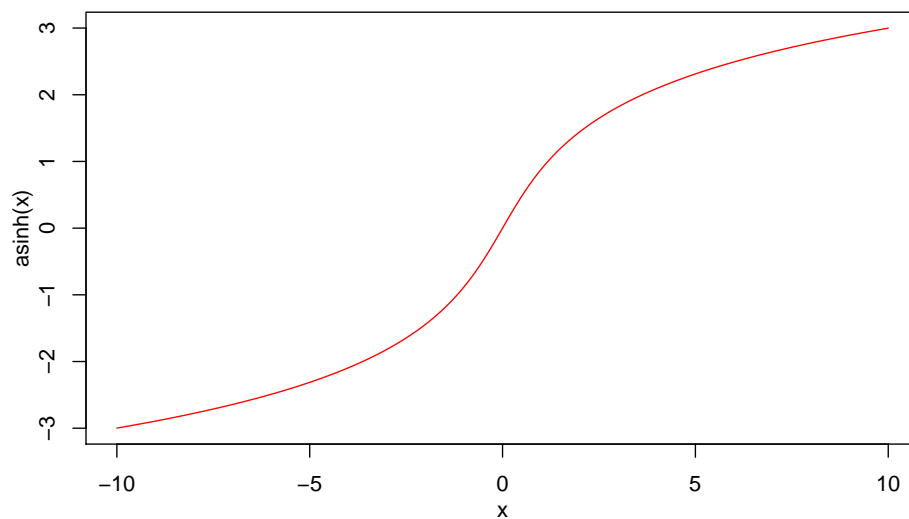
### 2.1 Elementary functions

Hyperbolic functions

$$\sinh x = \frac{e^x - e^{-x}}{2} \quad (2.1.1)$$



$$\operatorname{arcsinh} x = \log(x + \sqrt{x^2 + 1}) \quad (2.1.2)$$



### 2.2 Algebraic Functions

Gamma Function  $\Gamma(\cdot)$

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \quad (2.2.1)$$

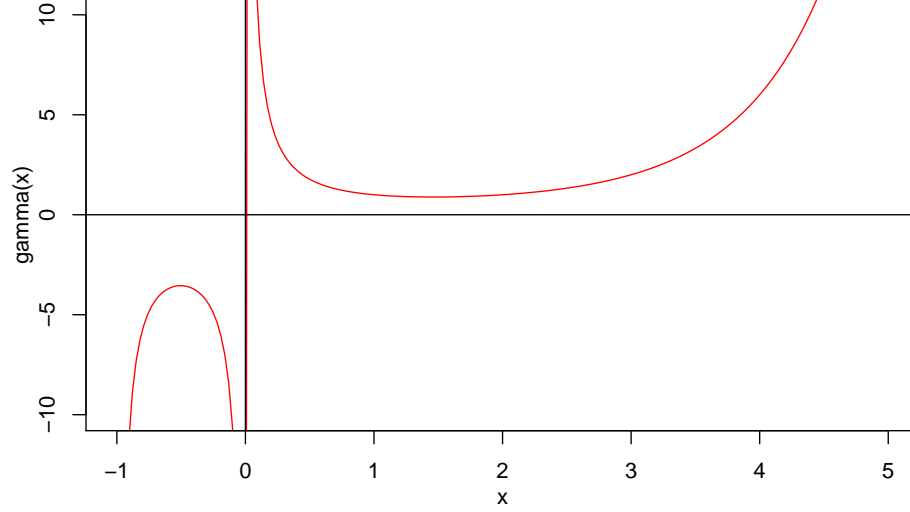
Properties:

$$\Gamma(1) = \Gamma(2) = 1 \quad (2.2.2)$$

$$\Gamma(n) = (n-1)! \quad (2.2.3)$$

$$\Gamma(n+1) = n\Gamma(n) \quad (2.2.4)$$

Proof: integrate by parts.



### Digamma function

$$\psi(x) = \frac{\partial}{\partial x} \log \Gamma(x) = \frac{1}{\Gamma(x)} \int_0^\infty t^{x-1} e^{-t} \log t \, dt. \quad (2.2.5)$$

Properties:

$$\psi(\alpha+1) = \frac{1}{\alpha} + \psi(\alpha) \quad (2.2.6)$$

$$\psi(1) = -0.5772 \quad \psi(2) = 0.4228 \quad (2.2.7)$$

### Multivariate gamma function

$$\Gamma_D(x) = \pi^{\frac{1}{4}D(D-1)} \prod_{i=1}^D \Gamma\left(x + \frac{1-i}{2}\right) \quad (2.2.8)$$

Note that  $\Gamma_1(x) = \Gamma(x)$ .

### Modified Bessel Function of the First Kind $I_\alpha(\cdot)$

$$I_\alpha(x) = \sum_{k=0}^{\infty} \frac{1}{k!(\alpha+k)!} \left(\frac{x}{2}\right)^{2k+\alpha} \quad (2.2.9)$$



**Beta function**  $B(a, b)$

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \int_0^1 x^{p-1}(1-x)^{q-1} dx, \quad (2.2.10)$$

where  $p > 0$  and  $q > 0$ .

## 2.3 Other functions

**Softmax function** For  $K$ -dimensional input vector  $\mathbf{x}$ , the  $c$ -th component of softmax is

$$S(\mathbf{x})_c = \frac{e^{x_c}}{\sum_k e^{x_k}}. \quad (2.3.1)$$

This describes discrete probability distribution between  $K$  states with corresponding weight  $e^{x_k}$ . In case of  $S(\mathbf{x}/T)$ , at the limit where  $T \rightarrow 0$ , all the mass is put on the state with the largest weight. If  $T \rightarrow \infty$ , all mass is spread uniformly across all the states.

**Softplus function** In 1-dimensional case

$$f(x) = \log(1 + e^x). \quad (2.3.2)$$

It is asymptotically equal to  $f(x) = 0$  if  $x \rightarrow -\infty$  and  $f(x) = x$  if  $x \rightarrow \infty$  and can be used as a smooth replacement for ReLU.

## 3 Calculus

### 3.1 Logarithm

Properties

$$\log x^\alpha = \alpha \log x \quad \text{and} \quad \log(xy) = \log x + \log y \quad (3.1.1)$$

### 3.2 Limits

#### 3.2.1 Limits in general

$\liminf$  and  $\limsup$  Definition:

$$\liminf \equiv \lim_{n \rightarrow \infty} \left( \inf_{m \geq n} x_m \right) \quad (3.2.1)$$

#### 3.2.2 Other limits

$$\lim_{x \rightarrow 0} x \cdot \log x = 0 \quad (3.2.2)$$

*Proof.* Write  $x \log x = (\log x)/(1/x)$  and use L'Hospital's rule.  $\square$

$$\lim_{x \rightarrow \infty} \frac{1}{x} \frac{\phi(x)}{1 - \Phi(x)} = 1, \quad (3.2.3)$$

where  $\phi(\cdot)$  and  $\Phi(\cdot)$  are normal density and distribution functions.

*Proof.* The fraction can be written as

$$\frac{\phi(x)}{1 - \Phi(x)} = \frac{\phi(x)}{\int_x^\infty \phi(s) ds}.$$

The integral can be expressed as the Rieman limit

$$\int_x^\infty \phi(s) ds = \lim_{\delta \rightarrow 0} [\phi(x)\delta + \phi(x + \delta)\delta + \phi(x + 2\delta)\delta + \dots].$$

Using the expression for  $\phi(\cdot)$  we get

$$\phi(x + \delta) = \phi(x)e^{-x\delta}e^{-\delta^2/2}$$

and hence we may write the denominator in (3.2.3) as

$$x\phi(x) \lim_{\delta \rightarrow 0} [1 + e^{-x\delta}e^{-\delta^2/2} + e^{-2x\delta}e^{-4\delta^2/2} + e^{-3x\delta}e^{-9\delta^2/2} + \dots]\delta.$$

This expressions  $e^{-x\delta}$ ,  $e^{-2x\delta}$  and so on for a geometric sequence with sum  $1/(1 - e^{-x\delta}) \approx 1/x\delta$  as  $x\delta \rightarrow 0$ . Accordingly, we let  $\delta \rightarrow 0$  and  $x \rightarrow \infty$  in such a way that  $x\delta \rightarrow 0$ . We have to show that the other terms  $e^{n^2\delta^2/2}$  do not “disturb” the geometric sequence too much.

Now find  $n^*$ , starting of which the residual sum on the geometric sequence  $1 + e^{-x\delta} + e^{-2x\delta} + \dots$  is less than  $\varepsilon > 0$ :

$$\sum_{n=n^*}^{\infty} e^{-nx\delta} < \varepsilon \quad \Rightarrow \quad n^* > -\frac{\log \varepsilon + \log(1 - e^{-x\delta})}{x\delta}$$

choose  $n^{**} > -\frac{\log \varepsilon}{x\delta} + 1 > n^*$ . Now find

$$\exp \frac{n^{**2}\delta^2}{2} = \exp \left( \frac{\log^2 \varepsilon}{2x^2} + \delta \frac{\log \varepsilon}{x} + \frac{\delta^2}{2} \right).$$

Because all the terms in parenthesis converge to as  $x \rightarrow \infty$  and  $\delta \rightarrow 0$ , the exponent converges to 1. Hence, at the limit, the Riemann sum is solely determined by the geometric sequence, and we have denominator in (3.2.3) equal to  $\phi(x)$ .  $\square$

### 3.2.3 Limits Involving Factorial

Stirling's formula:

$$\log n! = n \log n - n + O(\log n) \quad (3.2.4)$$

faktooriaalide jagatis

$$\lim_{n \rightarrow \infty} \frac{n!}{(n-m)!} = n^m \quad (3.2.5)$$

Märkus: siin on eeldatud, et  $m \nrightarrow \infty$ . Jaga läbi, arvesta, et  $n-1 \approx n$ . A special case:

$$\lim_{n \rightarrow \infty} \binom{n}{m} = \lim_{n \rightarrow \infty} \frac{n!}{(n-m)!m!} = \frac{n^m}{m!}. \quad (3.2.6)$$

## 3.3 Boundedness and Convergence

**Uniform Boundedness** Let  $X$  be a set and  $Y$  a metric space with metric  $d(\cdot, \cdot)$ . A family of functions  $\mathcal{F} = \{f_i : X \rightarrow Y, i \in I\}$  is uniformly bounded if there exist  $a \in Y$  and  $M \in \mathbb{R}$  such that

$$d(f_i(x), a) \leq M \quad \forall i \in I, x \in X \quad (3.3.1)$$

In particular,  $M$  and  $a$  must be a common elements for each  $i$ .

**Lipschitz Condition** (also *Lipschitz continuity*) Function  $f(x)$  satisfies Lipschitz condition if exists  $L > 0$  such that

$$|f(x_1) - f(x_2)| \leq L \cdot d(x_1, x_2) \quad \forall x_1, x_2 \text{ in range of } f(\cdot) \quad (3.3.2)$$

where  $d(\cdot, \cdot)$  is the metric distance.

**Equicontinuity**

**Pointwise equicontinuity** Let  $X$  and  $Y$  be metric spaces with metric  $d(\cdot, \cdot)$ . A family of functions  $\mathcal{F} = \{f_i : X \rightarrow Y, i \in I\}$  is equicontinuous at point  $x$  if for every  $\epsilon > 0$  there exists  $\delta(\epsilon, x)$  such that

$$\sup_{i \in I} d(f_i(\tilde{x}), f_i(x)) \leq \epsilon \quad \forall \tilde{x} : d(\tilde{x}, x) < \delta(\epsilon, x). \quad (3.3.3)$$

Note:  $\delta$  must not depend on  $i$ , but may depend on  $x, \epsilon$ .

## 3.4 Differentiation

### 3.4.1 Simple Derivatives

$$\frac{\partial}{\partial x} a^x = a^x \log a \quad (3.4.1)$$

$$\tan \phi' = \frac{1}{\cos^2 \phi} \quad (3.4.2)$$

$$\arcsin x' = \frac{1}{\sqrt{1-x^2}} \quad (3.4.3)$$

$$\arctan x' = \frac{1}{1+x^2} \quad (3.4.4)$$

### 3.4.2 Directional Derivative

$$f'(x; u) = \lim_{h \rightarrow 0} \frac{f(x + hu) - f(x)}{h} \quad (3.4.5)$$

### 3.4.3 Subderivative

For a convex function, *subderivative*  $f : I \rightarrow \mathbb{R}$  at a point  $x_0$  in the open interval  $I$  is a real number  $c$  such that

$$f(x) - f(x_0) \geq c(x - x_0) \quad (3.4.6)$$

for all  $x$  in  $I$ .

### 3.4.4 Jacobian Matrix

$$\mathcal{J} = \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} & \cdots & \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \frac{\partial(f_1, \dots, f_m)}{\partial(x_1, \dots, x_n)} = \mathbf{D} f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \quad (3.4.7)$$

### 3.4.5 Ümbrikuteoreem (*envelope theorem*)

Olgu  $M$  defineeritud kui optimum funktsioonist  $f$ :

$$M(a) = \max_x f(x, a), \quad (3.4.8)$$

kus  $a$  on parameeter. Siis

$$\frac{dM(a)}{da} = \left. \frac{\partial f(x^*, a)}{\partial a} \right|_{x^*=x(a)}. \quad (3.4.9)$$

### 3.4.6 Normal Density Related Derivatives

#### One-Dimensional Case

$$\frac{d}{dx}\Phi(-x) = \phi(x) \quad (3.4.10)$$

$$\frac{d}{dx}\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) = -\frac{1}{\sigma}\left(\frac{x-\mu}{\sigma}\right)\phi\left(\frac{x-\mu}{\sigma}\right) \quad (3.4.11)$$

$$\frac{d}{d\sigma}\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\left[\left(\frac{x-\mu}{\sigma}\right)^2 - 1\right]\phi\left(\frac{x-\mu}{\sigma}\right) \quad (3.4.12)$$

**Multi-Dimensional Case** Let  $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ ,  $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$  be the 2-dimensional variance-covariance matrix and  $\phi(\cdot, \cdot)$  the 2-D normal density, defined in (5.7.6):

$$\frac{\partial}{\partial x_1}\phi(\mathbf{x}, \Sigma) = \phi(\mathbf{x}, \Sigma)\frac{\rho x_2 - x_1}{1 - \rho^2} \quad (3.4.13)$$

$$\begin{aligned} \frac{\partial}{\partial \rho}\phi(\mathbf{x}, \Sigma) &= \phi(\mathbf{x}, \Sigma)\left[\frac{\rho}{1 - \rho^2}(1 - \mathbf{x}'\Sigma^{-1}\mathbf{x}) + \frac{x_1 x_2}{1 - \rho^2}\right] = \\ &= \phi(\mathbf{x}, \Sigma)\left[\frac{\rho}{1 - \rho^2} - \frac{\rho}{(1 - \rho^2)^2}(x_1^2 - 2\rho x_1 x_2 + x_2^2) + \frac{x_1 x_2}{1 - \rho^2}\right] \end{aligned} \quad (3.4.14)$$

$$\frac{\partial^2}{\partial x_1 \partial x_2}\phi(\mathbf{x}, \Sigma) = \phi(\mathbf{x}, \Sigma)\left[\frac{(x_1 - \rho x_2)(x_2 - \rho x_1)}{(1 - \rho^2)^2} + \frac{\rho}{1 - \rho^2}\right] \quad (3.4.15)$$

#### 3.4.7 Derivatives of Gamma Function

$$\Gamma'(x) = \psi(x)\Gamma(x) \quad (3.4.16)$$

$$\Gamma''(x) = \Gamma(x)\left[\psi'(x) + \psi^2(x)\right], \quad (3.4.17)$$

where  $\psi(x)$  is the [digamma function](#).

#### 3.4.8 Differentiation of Sums

$$\frac{\partial}{\partial x_i}\sum_j x_j = 1 \quad (3.4.18)$$

$$\frac{\partial}{\partial x_i}\left(\sum_j x_j\right)^2 = 2\sum_j x_j \quad (3.4.19)$$

#### 3.4.9 General Differentiation Rules

##### Pöördfunktsiooni tuletis

$$\frac{d}{dx}f^{-1}(x) = \frac{1}{\frac{d}{dx}f(x)}\bigg|_{x=f^{-1}(y)} = \frac{1}{f'[f^{-1}(y)]} \quad (3.4.20)$$

**Mitmekihilise (liit-) funktsiooni tuletis** Olgu  $v = v(b)$  ja  $f(v) = f(v(b)) = g(b)$ . Tuletised:

$$\frac{\partial g}{\partial b} = \frac{\partial f}{\partial v} \cdot \frac{\partial v}{\partial b} \equiv f'(v(b)) \cdot v'(b) \quad (3.4.21)$$

$$\frac{\partial^2 g}{\partial b^2} = \frac{\partial^2 f}{\partial v^2} \left( \frac{\partial v}{\partial b} \right)^2 + \frac{\partial f}{\partial v} \frac{\partial^2 v}{\partial b^2} \quad (3.4.22)$$

## 3.5 Integration

### 3.5.1 Simple Integrals

$$\int x^\alpha = \frac{1}{\alpha+1} x^{\alpha+1} + C, \quad \alpha \neq -1$$

$$\int \log x = x \log x - x + C \quad \text{proof: integrate by parts}$$

#### Leibnitz' Rule

$$\frac{\partial}{\partial x} \int_{a(x)}^{b(x)} f(y) \, dy = f[b(x)]b'(x) - f[a(x)]a'(x) \quad (3.5.1)$$

$$\begin{aligned} \frac{\partial}{\partial x} \int_{a(x)}^{b(x)} f(y, x) \, dy &= f[b(x), x]b'(x) - f[a(x), x]a'(x) + \\ &+ \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(y, x) \, dy \end{aligned} \quad (3.5.2)$$

### 3.5.2 Integrals related to probability distributions

**Normal distribution** Let

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}. \quad (3.5.3)$$

To prove the following, simply integrate  $\phi(\cdot)$ , in most cases by parts.

$$\int \phi(x) \, dx \equiv \Phi(x) \quad (3.5.4)$$

$$\int_a^b \phi\left(\frac{x-\mu}{\sigma}\right) \, dx = \sigma \Phi\left(\frac{b-\mu}{\sigma}\right) - \sigma \Phi\left(\frac{a-\mu}{\sigma}\right) \quad (3.5.5)$$

$$\int \phi^2(x) \, dx = \frac{1}{2\pi} \Phi(\sqrt{2}x) \quad (3.5.6)$$

$$\int x\phi(x) \, dx = -\phi(x) \quad (3.5.7)$$

$$\int_{-\infty}^{\infty} x\phi\left(\frac{x-\mu}{\sigma}\right) \, dx = \sigma\mu \quad (3.5.8)$$

$$\int x\phi\left(\frac{x-\mu}{\sigma}\right) \, dx = \sigma\mu\Phi\left(\frac{x-\mu}{\sigma}\right) - \sigma^2\phi\left(\frac{x-\mu}{\sigma}\right) + C \quad (3.5.9)$$

$$\begin{aligned} \int_a^b x\phi\left(\frac{x-\mu}{\sigma}\right)dx &= \sigma\mu\left[\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)\right] + \\ &+ \sigma^2\left[\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)\right] \end{aligned} \quad (3.5.10)$$

$$\int x^2\phi(x)dx = -x\phi(x) + \Phi(x) \quad (3.5.11)$$

$$\int_a^b x^2\phi(x)dx = a\phi(a) - b\phi(b) + \Phi(b) - \Phi(a) \quad (3.5.12)$$

$$\int x^2\phi\left(\frac{x}{\sigma}\right)dx = -\sigma^2x\phi\left(\frac{x}{\sigma}\right) + \sigma^3\Phi\left(\frac{x}{\sigma}\right) \quad (3.5.13)$$

$$\int_{-\infty}^{\infty} x^2\phi\left(\frac{x-\mu}{\sigma}\right)dx = \sigma(\mu^2 + \sigma^2) \quad (3.5.14)$$

$$\int_{-\infty}^{\infty} e^{ax}\phi\left(\frac{x-\mu}{\sigma}\right)dx = e^{\alpha\mu + \frac{1}{2}\sigma^2\alpha^2} \quad (3.5.15)$$

$$\int_s^t \phi(u) \log \phi(u) du = \frac{1}{2} [\Phi(s) - \Phi(t)] (1 + \log 2\pi) + \frac{1}{2} [t\phi(t) - s\phi(s)] \quad (3.5.16)$$

$$\begin{aligned} \int_s^t \phi(u) \log \phi(u - \mu) du &= \frac{1}{2} [\Phi(s) - \Phi(t)] (1 + \log 2\pi + \mu^2) + \\ &+ \frac{1}{2} [t\phi(t) - s\phi(s)] + \mu [\phi(s) - \phi(t)] \end{aligned} \quad (3.5.17)$$

$$\int x\phi^2(x)dx = -\frac{1}{2}\phi^2(x) \quad (3.5.18)$$

$$\int \phi(x)\Phi(x)dx = \frac{1}{2}\Phi^2(x) \quad (3.5.19)$$

$$\int x\phi(x)\Phi(x)dx = -\phi(x)\Phi(x) + \frac{1}{2\sqrt{\pi}}\Phi(\sqrt{2}x) \quad (3.5.20)$$

$$\int x^2\phi(x)\Phi(x)dx = \frac{1}{2}\Phi^2(x) - x\phi(x)\Phi(x) - \frac{1}{2}\phi^2(x) \quad (3.5.21)$$

**Log-normal density** Let  $f(\cdot)$  be the log-normal density.

$$\begin{aligned} \int_a^\infty xf(x)dx &= \int_a^b \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right] dx = \\ &= \frac{1 - \Phi\left(\frac{\log a - \mu - \sigma^2}{\sigma}\right)}{1 - \Phi\left(\frac{\log a - \mu}{\sigma}\right)} e^{\mu + \frac{1}{2}\sigma^2} \end{aligned} \quad (3.5.22)$$

$$\begin{aligned} \int_a^b xf(x)dx &= \int_a^b \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right] dx = \\ &= e^{\mu + \frac{1}{2}\sigma^2} \left[ -\Phi\left(\frac{\log a - \mu - \sigma^2}{\sigma}\right) + \Phi\left(\frac{\log b - \mu - \sigma^2}{\sigma}\right) \right] \end{aligned} \quad (3.5.23)$$



### 3.5.3 Other integrals

$$\int_0^t e^{-rT} dT = \frac{1}{r} [1 - e^{-rt}] \quad (3.5.24)$$

$$\int \frac{dx}{(p + qe^{ax})^2} = \frac{x}{p^2} + \frac{1}{ap(p + qe^{ax})} - \frac{1}{ap^2} \ln(p + qe^{ax}) \quad (3.5.25)$$

$$\int \log x \, dx = x \log x - x \quad (3.5.26)$$

$$\int e^{-ax} dx = \frac{1}{a} [1 - e^{-ax}] \quad (3.5.27)$$

$$\int x e^x dx = x e^x - e^x \quad (3.5.28)$$

$$\int x e^{-x} dx = -x e^{-x} - e^{-x} \quad (3.5.29)$$

$$\int x e^{ax} dx = \frac{x}{a} e^{ax} - \frac{1}{a^2} e^{ax} \quad (3.5.30)$$

$$\int_a^b x e^{ax} dx = \frac{1}{a} \left[ e^{ab} \left( b - \frac{1}{a} \right) - e^{aa} \left( a - \frac{1}{a} \right) \right] \quad (3.5.31)$$

$$\int x^2 e^{ax} dx = \frac{1}{a} x^2 e^{ax} - \frac{2}{a} \int x e^{ax} dx \quad (3.5.32)$$

$$\int_a^b x^2 e^{ax} dx = \frac{1}{a} \left[ e^{ab} \left( b^2 - \frac{2b}{a} + \frac{2}{a^2} \right) - e^{aa} \left( a^2 - \frac{2a}{a} + \frac{2}{a^2} \right) \right] \quad (3.5.33)$$

$$\int_0^\infty x^\alpha e^{-\beta x} dx = \frac{\Gamma(\alpha + 1)}{\beta^{\alpha+1}} \quad (3.5.34)$$

Let  $f(\cdot)$  be a distribution function and  $\bar{F}(\cdot)$  the corresponding survival function:

$$\int_c^b \left[ f(x) \int_c^x w(y) dy \right] dx = \int_c^b \bar{F}(x) w(x) dx \quad (3.5.35)$$

### Euleri konstant

$$\int_0^\infty e^{-z} \log z \, dz = c \approx -0,5772 \quad (3.5.36)$$

### 3.5.4 Üldised integreerimise reeglid

**Muutuja vahetus integraali all** Olgu vaja üle minna muutujatelt  $(x_1, \dots, x_N)$  muutujatele  $(y_1, \dots, y_N)$ . Sel juhul

$$\begin{aligned} \int_V f(x_1, \dots, x_N) dx_1 \dots dx_N &= \\ &= \int_V f[y_1(x_1, \dots, x_N), \dots, y_N(x_1, \dots, x_N)] \frac{dy_1 \dots dy_N}{|\mathcal{J}|} = \\ &= \int_V g(y_1, \dots, y_N) \frac{dy_1 \dots dy_N}{|\mathcal{J}|}, \end{aligned} \quad (3.5.37)$$

kus

$$|\mathcal{J}| = \left| \frac{\partial(y_1, \dots, y_N)}{\partial(x_1, \dots, x_N)} \right| \quad (3.5.38)$$

on koordinaatteisenduse jakobjaani absoluutväärtus.

### 3.5.5 Laplace'i teisendus

Laplace'i teisendus juhusliku muutuja  $X$  jaotusfunktsioonist on

$$L_f(s) = \mathbb{E} e^{-sX} = \int e^{-sx} dF_X(x). \quad (3.5.39)$$

Laplace'i teisendus on sama mis momendifunktsioon.

### 3.5.6 Numbriline integreerimine

**Monte-Carlo integraal** Olgu vaja leida

$$I = \int_a^b f(x) dx = (b-a) \int_a^b f(x) \frac{1}{b-a} dx = (b-a) \mathbb{E}[f(X)], \quad (3.5.40)$$

kus  $X \sim U(a, b)$ . Valimis suurusega  $N$  olgu  $x_1, \dots, x_N \sim i.i.d U(a, b)$ . Siis integraali hinnanguks on funktsiooni väärtuste keskmine ja veahinnanguks tema standardhälve valimis:

$$\hat{I}_N = (b-a) \frac{1}{N} \sum_{i=1}^N f(x_i) \quad (3.5.41)$$

$$\widehat{\text{Var}} \hat{I}_N = \frac{(b-a)^2}{N} \frac{1}{N} \sum_{i=1}^N \left[ f(x_i) - \frac{1}{N} \hat{I}_N \right]^2 \quad (3.5.42)$$

### 3.6 Differential Equations

Let  $c$  be a constant.

$$\begin{aligned}y(x)' + Py &= Q \\y(x) &= \frac{Q}{P} + ce^{-Px}\end{aligned}\tag{3.6.1}$$

$$\begin{aligned}y(x)' + P(x)y &= Q(x) \\y(x) &= e^{-\int P(x) dx} \int Q(x)e^{\int P(x) dx} dx + ce^{-\int P(x) dx}\end{aligned}\tag{3.6.2}$$

### 3.7 Optimization

#### 3.7.1 Second-order maximum/minimum conditions for constrained optimisation

The problem is

$$\begin{aligned} \max z &= f(x_1, x_2, \dots, x_n) \\ \text{s.t. } g(x_1, x_2, \dots, x_n) &= 0. \end{aligned} \quad (3.7.1)$$

Corresponding Lagrangian is

$$Z = f(x_1, x_2, \dots, x_n) - \lambda g(x_1, x_2, \dots, x_n). \quad (3.7.2)$$

Corresponding *bordered Hessian* is:

$$|\bar{H}| = \begin{vmatrix} 0 & g_1 & g_2 & \dots & g_n \\ g_1 & Z_{11} & Z_{12} & \dots & Z_{1n} \\ g_2 & Z_{21} & Z_{22} & \dots & Z_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_n & Z_{n1} & Z_{n2} & \dots & Z_{nn} \end{vmatrix} \quad (3.7.3)$$

and successive *principal minors* are:

$$|\bar{H}_2| = \begin{vmatrix} 0 & g_1 & g_2 \\ g_1 & Z_{11} & Z_{12} \\ g_2 & Z_{21} & Z_{22} \end{vmatrix} \quad |\bar{H}_3| = \begin{vmatrix} 0 & g_1 & g_2 & g_3 \\ g_1 & Z_{11} & Z_{12} & Z_{13} \\ g_2 & Z_{21} & Z_{22} & Z_{23} \\ g_3 & Z_{31} & Z_{32} & Z_{33} \end{vmatrix} \quad \dots \quad |\bar{H}_n| = |\bar{H}| \quad (3.7.4)$$

The second derivative  $d^2z$  is positive definite iff

$$\bar{H}_2 < 0, \quad \bar{H}_3 < 0, \quad \dots, \quad \bar{H}_n < 0,$$

and negative definite iff

$$\bar{H}_2 > 0, \quad \bar{H}_3 < 0, \quad \bar{H}_n > 0, \quad \dots$$

Note that  $\bar{H}_1$  is always negative.

Proof: [Chiang \(1984\)](#).

### 3.7.2 Optimal control (dynamic optimisation)

The problem is:

$$\begin{aligned} & \max_u \int_0^T F(t, y, u) dt \\ & \text{s.t. } \dot{y} = f(t, y, u) \\ & y(0) = A \quad y(T) \text{ free.} \end{aligned} \quad (3.7.5)$$

Corresponding *Hamiltonian* is

$$H(t, y, u, \lambda) \equiv F(t, y, u) + \lambda(t)f(t, y, u). \quad (3.7.6)$$

The first order conditions for optimum are

1.  $\max_u H(t, y, u, \lambda) \quad \forall t \in [0, T]$  or, less generally,  $\frac{\partial H}{\partial u} = 0$  (optimality condition).
2.  $\dot{y} = \frac{\partial H}{\partial \lambda}$  (equation of motion for  $y$ ).
3.  $\dot{\lambda} = -\frac{\partial H}{\partial y}$  (equation of motion for  $\lambda$ ).
4.  $\lambda(T) = 0$  (transversality condition).

Proof: [Miller \(1979\)](#)

### 3.7.3 Newton-Raphson maximization

Non-linear continuous function of  $N$ -dimensional parameter can, under suitable assumptions, be approximated as  $N$ -dimensional parabola. When running non-linear maximization, we may approximate the function in this way at the initial value of the parameter vector. The maximum of the approximation can be used as the initial value for the next step.

Let us maximise a function  $l(\boldsymbol{\vartheta})$  where  $\boldsymbol{\vartheta}$  is a  $N$ -dimensional parameter vector. Let the initial value of the parameter be  $\boldsymbol{\vartheta}_0$ . From Taylor's approximation:

$$l(\boldsymbol{\vartheta}) \approx l(\boldsymbol{\vartheta}_0) + \left. \frac{\partial l(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) + \frac{1}{2} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)' \left. \frac{\partial^2 l(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} (\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0) \quad (3.7.7)$$

At the maximum  $\partial l(\boldsymbol{\vartheta}) / \partial \boldsymbol{\vartheta} = 0$  and hence the parameter value at the maximum (the initial value for the next iteration):

$$\boldsymbol{\vartheta}_1 = \boldsymbol{\vartheta}_0 - \left[ \left. \frac{\partial^2 l(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \right]^{-1} \left. \frac{\partial l(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \right|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \quad (3.7.8)$$

The algorithm requires either programming the analytical Hessian matrix  $\frac{\partial^2 l(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'}$ , or calculating the Hessian matrix by numeric differentiation. The first way may be complicated, the latter one slow and subject to numerical errors.

**BHHH maximization** BHHH is a particular version of Newton-Raphson algorithm, suitable for maximizing log-likelihood function only. BHHH uses the information equality for approximating the Hessian:

$$\mathbb{E} \left[ \frac{\partial^2 l(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right]_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} = - \mathbb{E} \left[ \frac{\partial l(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}'} \bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \frac{\partial l(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \bigg|_{\boldsymbol{\vartheta}=\boldsymbol{\vartheta}_0} \right] \quad (3.7.9)$$

This algorithm does not require Hessian matrix (this is approximated). However, it typically requires around 10 times more iterations for convergence as the approximation may be quite imprecise when initial values are far off the target. Note also that while the estimates are exactly the same as in the case of NR algorithm, the standard errors may be different on a finite sample ([Calzolari and Fiorentini, 1993](#)).

## 4 Algebra

### 4.1 Mõisted

**proper subset**  $A$  on  $B$  *proper subset* kui  $A \subseteq B$  kui  $B \not\subseteq A$ .

**proper subspace** Kui  $A$  ja  $B$  ja  $A$  on  $B$  *proper subset*. Näiteks tasandi tõeline (lineaarne) alamruum on sirge.

### 4.2 Sets

#### 4.2.1 Definitions

**Lower Bound**  $a$  is lower bound of  $S$  if  $a \leq x$  for all  $x \in S$ .

**Infimum**

$$a = \inf S \Leftrightarrow a \text{ is the largest lower bound of } S \quad (4.2.1)$$

#### 4.2.2 Basic Operations

**De Morgan's Laws**

$$\left( \bigcap_{A \in F} A \right)^C = \bigcup_{A \in F} A^C \quad (4.2.2)$$

$$\left( \bigcup_{A \in F} A \right)^C = \bigcap_{A \in F} A^C \quad (4.2.3)$$

**Other**

$$A \setminus B = A \cap B^C \quad (4.2.4)$$

$$\left( \bigcup_i A_i \right) \setminus \left( \bigcup_i B_i \right) \subseteq \bigcup_i (A_i \setminus B_i) \quad (4.2.5)$$

**Sigma Field** A set  $S$  is *sigma field* iff

1.  $\mathbb{R} \in S$
2.  $S$  is closed under complement:  $E \in S \Rightarrow E^C \in S \quad \forall E \in S$
3.  $S$  is closed under countable union:  $\bigcup_i^n E_i \in S \quad E_i \in S, i = 1, 2, \dots$
4.  $S$  is closed under countable intersection:  $\bigcap_i^n E_i \in S \quad E_i \in S, i = 1, 2, \dots$   
This follows from 2,3 and De Morgan's laws.

### 4.3 Simple Algebra

#### Binomial Theorem

$$(a + b)^n = \sum_{i=0}^n \binom{n}{i} a^{n-i} b^i \quad (4.3.1)$$

where  $\binom{n}{i}$  is the *binomial coefficient*, number of distinct combinations of  $i$  elements out of  $n$  elements in total:  $\binom{n}{i} \equiv C_n^i = \frac{n!}{(n-i)!i!}$ .

$$\sum_{i=0}^n \binom{n}{i} e^{\alpha i} = (1 + e^\alpha)^n \quad (4.3.2)$$

$$\sum_{i=0}^n i \binom{n}{i} e^{\alpha i} = n e^\alpha (1 + e^\alpha)^{n-1} \quad (4.3.3)$$

#### Geomeetrilise jada summa

$$S = 1 + q + q^2 + q^3 + \dots = \frac{1}{1-q}. \quad (4.3.4)$$

Tõestus: kirjuta välja  $qS$ , lahuta ja avalda  $S$ . Märkus: kui jada on kujul  $S' = q + q^2 + q^3 + \dots$ , siis  $S' = qS$ . Oluline erijuht kui  $q = \frac{1}{1+r}$ :

$$S = 1 + \frac{1}{1+r} + \frac{1}{(1+r)^2} + \dots = 1 + \frac{1}{r}. \quad (4.3.5)$$

$$\sum_{i=0}^{\infty} i p^i = \frac{p}{(1-p)^2} \quad (4.3.6)$$

Tõestus: kui  $S$  on antud summa, siis avalda  $S - pS \dots$

#### EkspONENT piirväärtusena

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a}{n}\right)^n = e^a. \quad (4.3.7)$$

Tõestus: arenda Newtoni binoomvalemiga ritta, arvesta (3.2.5).

### 4.4 Taylori rida

**Taylori rida** Iga funktsiooni võib punkti  $x_0$  ümbruses esitada astmereana:

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + \dots = \\ &= \sum_{i=0}^{\infty} f^{(i)}(x_0) \frac{(x - x_0)^i}{i!}. \end{aligned} \quad (4.4.1)$$

Tõestus: kirjuta samasugune astmerida tundmatute kordajatega välja, võrruta  $f(x - x_0)$ -ga ja võta järjest tuletisi. Taylori rea erijuht, kui  $x_0 = 0$  on McLaureni rida:

$$f(x) = f(0) + f'(0)x + \frac{1}{2}f''(0)x^2 + \dots = \sum_{i=0}^{\infty} f^{(i)}(0) \frac{x^i}{i!}. \quad (4.4.2)$$



### **Eksponeendi astmerida**

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^3 + \dots = \sum_{i=0}^{\infty} \frac{x^i}{i!}. \quad (4.4.3)$$

Tõestus: arenda  $e^x$  Tayloriga ritta.

### **Eksponeendi piirväärtus**

$$\lim_{x \rightarrow 0} e^x = 1 + x. \quad (4.4.4)$$

Tõestus: eksponeendi astmerekord. Märkus: piirväärtus  $1 + x$  on kõige tavalisem, mida on vaja kasutada. Olenevalt ülesandest tuleb arvestada rohkem (või ka vähem) astmerekord liikmeid.

### **Logaritmi astmerida**

$$\log x = \log x_0 + \frac{1}{x_0}(x - x_0) - \frac{1}{2x_0^2}(x - x_0)^2 + \frac{1}{3x_0^3}(x - x_0)^3 - \dots \quad (4.4.5)$$

$$= (x - 1) - \frac{(x - 1)^2}{2} + \frac{(x - 1)^3}{3} - \frac{(x - 1)^4}{4} + \dots \quad (4.4.6)$$

Tõestus: arenda Tayloriga ritta  $x_0 = 1$  ümbruses.

## 4.5 Summade astmed

$$\left( \sum_{i=1}^N x_i \right)^2 = \sum_{i=1}^N x_i^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^N x_i x_j \quad (4.5.1)$$

$$\left( \sum_{i=1}^N x_i \right)^3 = \sum_{i=1}^N x_i^3 + 3 \sum_{\substack{i,j=1 \\ i \neq j}}^N x_i x_j^2 + \sum_{\substack{i,j,k=1 \\ i \neq j; j \neq k; k \neq i}}^N x_i x_j x_k \quad (4.5.2)$$

$$\begin{aligned} \left( \sum_{i=1}^N x_i \right)^4 &= \sum_{i=1}^N x_i^4 + 4 \sum_{\substack{i,j=1 \\ i \neq j}}^N x_i x_j^3 + 3 \sum_{\substack{i,j=1 \\ i \neq j}}^N x_i^2 x_j^2 + 6 \sum_{\substack{i,j,k=1 \\ i \neq j; j \neq k; k \neq i}}^N x_i x_j x_k^2 + \\ &+ \sum_{\substack{i,j,k,l=1 \\ i,j,k,l \neq}}^N x_i x_j x_k x_l \end{aligned} \quad (4.5.3)$$

Ühekordsetes summates on  $N$  liiget, kahekordsetes  $N(N-1)$ , kolmekordsetes  $N(N-1)(N-2)$  ning neljakordsetes  $N(N-1)(N-2)(N-3)$ .

Tuletuskäik lähtub viimasel juhul niisugustest mõtetest:

- Kui komponentide indeksid ei tohi olla võrdsed, siis on järgmist komponenti võimalik valida ühe võrra vähem
- Esimesel liikmel võetakse sisse kõik komponendid, seega on  $C_0^4$  varianti.
- Teisel liikmel on kaks komponenti ( $x_i$  ja  $x_j$ ), ühte võetakse kolm korda, teist korra. Seega tuleb valida üks, mis iga kord välja jäetakse. Seega  $C_1^4$  võimalust.
- Kolmandal liikmel valitakse mõlemad komponendid kahe kaupa. Kokku on  $C_2^4 = 6$  võimalust kahe kaupa valida, kuna aga pole vahet kumb komponentidest on kumb, siis jääb järele pool nendest.
- Neljandal liikmel on kolm komponenti, ruutliikme valimiseks on  $C_2^4 = 6$  võimalust. Kuna teised liikmed on esimeses astmes, siis on küll vahe, kumbad me välja valime. Jääb 6.
- Viimane, kõiki üks kord,  $C_4^4 = 1$ .

Kui  $X \sim i.i.d$ , siis

$$\mathbb{E} \left( \sum_{i=1}^N x_i \right)^2 = N \mathbb{E} X^2 + N(N-1)(\mathbb{E} X)^2 = N^2(\mathbb{E} X)^2 + N \text{Var } X \quad (4.5.4)$$

## 4.6 Linear Algebra

### 4.6.1 Matrix

**Toeplitz matrix** (diagonal-constant matrix) is a matrix where the elements are constant along the diagonals:

$$M = \begin{bmatrix} a_0 & a_{-1} & a_{-2} & \cdots & \cdots & a_{-(n-1)} \\ a_1 & a_0 & a_{-1} & \ddots & & \vdots \\ a_2 & a_1 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & a_{-1} & a_{-2} \\ \vdots & & \ddots & a_1 & a_0 & a_{-1} \\ a_{n-1} & \cdots & \cdots & a_2 & a_1 & a_0 \end{bmatrix}$$

### 4.6.2 Determinant

Let

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad (4.6.1)$$

Now

$$|M| = |A - BD^{-1}C| \cdot |D| \quad (4.6.2)$$

### 4.6.3 Inverse Matrix

**Inverse matrix**

$$A^{-1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{|A|} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \quad (4.6.3)$$

Partitioned inverse formula

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} E^{-1} & -E^{-1}BD^{-1} \\ -D^{-1}CE^{-1} & F^{-1} \end{bmatrix} \quad (4.6.4)$$

where

$$E = A - BD^{-1}C \equiv M/D \quad \text{is the Schur complement of } M \text{ wrt } D$$

$$E^{-1} = A^{-1} + A^{-1}BF^{-1}CA^{-1}$$

$$F = D - CA^{-1}B \equiv M/A$$

$$F^{-1} = D^{-1} + D^{-1}CE^{-1}BD^{-1}$$

Proof: [Murphy \(2012, pp 118-119\)](#)

**Moore-Penrose pseudoinverse**

$$A^+ = \lim_{\alpha \downarrow 0} (A^T A + \alpha I)^{-1} A^T \quad (4.6.5)$$

M-P pseudoinverse solves the equation

$$\mathbf{y} = \mathbf{A}\mathbf{x} \Rightarrow \mathbf{x} = \mathbf{A}^+ \mathbf{y} \quad (4.6.6)$$

in general case. If  $\mathbf{A}$  has more columns than rows (infinite number of solutions), it picks the one that has minimal Euclidean norm of  $\mathbf{x}$ . If  $\mathbf{A}$  has more rows than columns (no solutions), it picks the one with minimal Euclidean norm of  $\mathbf{y} - \mathbf{A}\mathbf{x}$ .

#### 4.6.4 Matrix norm

**Frobenius norm** *Frobenius norm* is an analogue of  $L_2$  norm for matrices:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{\text{Tr}(\mathbf{A} \cdot \mathbf{A}^T)} \quad (4.6.7)$$

### 4.7 Differentiation of Matrices and Vectors

Definition: let  $\lambda$  be a scalar,  $\mathbf{x}$  and  $\boldsymbol{\beta}$   $K \times 1$  vectors, and  $\mathbf{A}$  a  $K \times K$  matrix. Define:

$$\frac{\partial \lambda}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1} \lambda \\ \vdots \\ \frac{\partial}{\partial x_K} \lambda \end{bmatrix} \quad (4.7.1)$$

More results:

$$\frac{\partial \mathbf{x}}{\partial \mathbf{x}'} = \begin{bmatrix} \frac{\partial x_1}{\partial x_1} & \frac{\partial x_1}{\partial x_2} & \cdots & \frac{\partial x_1}{\partial x_K} \\ \frac{\partial x_2}{\partial x_1} & \frac{\partial x_2}{\partial x_2} & \cdots & \frac{\partial x_2}{\partial x_K} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_K}{\partial x_1} & \frac{\partial x_K}{\partial x_2} & \cdots & \frac{\partial x_K}{\partial x_K} \end{bmatrix} = \mathbf{I} \quad \frac{\partial \mathbf{x}'}{\partial \mathbf{x}} = \mathbf{I} \quad (4.7.2)$$

$$\frac{\partial \mathbf{x}' \boldsymbol{\beta}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1} (\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K) \\ \frac{\partial}{\partial x_2} (\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K) \\ \vdots \\ \frac{\partial}{\partial x_K} (\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K) \end{bmatrix} = \boldsymbol{\beta} \quad (4.7.3)$$

$$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}'} = \mathbf{A} \quad \frac{\partial \mathbf{x}' \mathbf{A}}{\partial \mathbf{x}} = \mathbf{A} \quad (4.7.4)$$

$$\frac{\partial \mathbf{x}' \mathbf{x}}{\partial \mathbf{x}} = \mathbf{x} + \mathbf{x}' = 2\mathbf{x} \quad \frac{\partial \mathbf{x}' \mathbf{A} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}')\mathbf{x} \quad (4.7.5)$$

$$\frac{\partial}{\partial \mathbf{A}} \text{Tr}(\mathbf{B}\mathbf{A}) = \mathbf{B}' \quad \frac{\partial}{\partial \mathbf{A}} \log |\mathbf{A}| = (\mathbf{A}^{-1})' \quad (4.7.6)$$

Let  $\mathbf{f}(\mathbf{x})$  be a  $M \times 1$  vector function and  $\mathbf{B}$  a  $M \times M$  matrix.

$$\frac{\partial \mathbf{f}(\mathbf{x})' \mathbf{B} \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} = [\mathbf{D} \mathbf{f}(\mathbf{x})]' (\mathbf{B} + \mathbf{B}') \mathbf{f}(\mathbf{x}), \quad (4.7.7)$$

where  $\mathbf{D} \mathbf{f}(\mathbf{x})$  is [Jacobian Matrix](#).

Oluline: ühte korrutamist ei tohi teiseks muuta. Näiteks kui avaldis sisaldab nii maatrikskorrutist kui skalaariga korrutamist (skalaariga korrutamine on põhimõtteliselt sama mis Kroneckeri korrutis  $\otimes$ ), ei tohi endist skalaariga korrutamist diferentseerimise järel tõlgendada maatrikskorrutisena. Mis siis et skalaari asemel on nüüd maatriks:

$$\frac{\partial}{\partial \beta'} [(\beta' x) \otimes y] = \frac{\partial \beta' x}{\partial \beta'} \otimes y = x' \otimes y = y x'. \quad (4.7.8)$$

Skalaariga korrutamisel korrutatakse kõik maatriksi elemendid läbi sama skalaariga, seega pääle tuletise võtmist tuleb kõik vektori  $y$  elemendid läbi korrutada tuletisvektoriga  $x'$ .

## 4.8 Võrratused

### 4.8.1 Hölder's inequality

Let  $X$  and  $Y$  be random variables.

$$\mathbb{E} |XY| \leq \left\{ \mathbb{E} \left[ |X|^{\frac{1}{\alpha}} \right] \right\}^{\alpha} \left\{ \mathbb{E} \left[ |Y|^{\frac{1}{1-\alpha}} \right] \right\}^{1-\alpha} \quad (4.8.1)$$

### 4.8.2 Jensen Inequality

Let  $f(\cdot)$  be a concave function:

$$\sum_i \lambda_i f(x_i) \leq f \left( \sum_i \lambda_i x_i \right), \quad \sum \lambda_i = 1 \quad (4.8.2)$$

$$\mathbb{E} f(x) < f(\mathbb{E} x). \quad (4.8.3)$$

Proof: definition of concavity, induction.

### 4.8.3 Triangle Inequality

$$|x + y| \leq |x| + |y| \quad (4.8.4)$$

### 4.8.4 Cauchy-Schwartzi võrratus

$$| \langle x, y \rangle | \leq \|x\| \cdot \|y\| \quad (4.8.5)$$

For sequences

$$\left( \sum a_i b_i \right)^2 \leq \left( \sum a_i^2 \right) \left( \sum b_i^2 \right). \quad (4.8.6)$$

For functions:

$$\langle x, y \rangle = \int x \cdot y \, dx \quad \|x\| = \sqrt{\langle x, x \rangle}$$

Vektorkujul:

$$(a \cdot b)^2 \leq \|a\|^2 \|b\|^2 \quad (4.8.7)$$

ehk siis ka

$$\sum_i z_i z'_i \geq \frac{\sum_i a_i z_i \sum_i a_i z'_i}{\sum_i a_i^2} \quad (4.8.8)$$

### 4.8.5 Inequalities, containing exponent

Proof in most cases by analysing the corresponding function.

$$e^a \geq a \quad (4.8.9)$$

$$1 - e^{-a} < a \quad (4.8.10)$$

$$(1 - e^{-a})e^{-a} < a \quad (4.8.11)$$

$$(1 + a)e^a \geq (1 + 2a) \quad (4.8.12)$$

$$(1 + a)e^{-a} < 1 \quad \text{if } a > 0 \quad (4.8.13)$$

$$(a - 1)e^a > -1 \quad \text{if } a > 0 \quad (4.8.14)$$

$$(1 + a^2)e^{-a} < 1 \quad \text{if } a > 0 \quad (4.8.15)$$

$$e^{-a} - e^{-b} < -a + b \quad \text{if } 0 < a < b \quad (4.8.16)$$

$$e^a - e^b = (a - b) + \frac{1}{2}(a^2 - b^2) + \frac{1}{6}(a^3 - b^3) + \dots \quad (4.8.17)$$

$$\geq (a - b) \quad \text{if } a \geq b \quad (4.8.18)$$

$$ae^{-b} - be^{-a} \geq a - b \quad \text{if } a \geq b \quad (4.8.19)$$

$$a^2e^b - b^2e^a \geq a^2 - b^2 + ab(a - b) + \frac{1}{6}a^2b^2(b - a) + \frac{1}{24}a^2b^2(b^2 - a^2) + \dots$$

$$\text{if } b \geq a \quad (4.8.20)$$

## 4.9 Differential Equations

### 4.9.1 Linear Equations with Constant Coefficients

**Homogeneous linear differential equations with constant coefficients** are of form

$$a_n y^{(n)} + a_{n-1} y^{(n-1)} + \dots + a_1 y = 0 \quad (4.9.1)$$

It's particular solution is in the form

$$y = e^{cx} \quad (4.9.2)$$

where  $c$  is a root of the characteristic polynomial

$$a_n c^n + a_{n-1} c^{n-1} + \dots + a_1 c = 0. \quad (4.9.3)$$

The general solution is

$$y(x) = u_1 e^{c_1 x} + u_2 e^{c_2 x} + \dots + u_n e^{c_n x} \quad (4.9.4)$$

where  $c_i, i = 1 \dots n$ , are the characteristics roots of (4.9.3) and  $u_i$  are constants to be determined.

Proof: insert  $y(x) = e^{cx}$  into (4.9.1) and take derivatives.

**First order linear differential equation** has a general form

$$y'(x) + p(x)y(x) = q(x) \quad (4.9.5)$$

It's solution is

$$y(x) = \frac{\int u(x)q(x)dx + C}{u(x)} \quad (4.9.6)$$

where the *integrating factor*,  $u(x)$ , is

$$u(x) = e^{\int p(x)dx} \quad (4.9.7)$$

## 5 Probability and Statistics

### 5.1 Combinatorics

**Combinations** Number of combinations of selecting  $k$  items out of  $n$  where order does not matter:

$$\binom{n}{k} \equiv C_k^n = \frac{n!}{k!(n-k)!} \quad (5.1.1)$$

### 5.2 Random Variables

#### 5.2.1 General Concepts

**Variance**

$$\text{Var } X = \mathbb{E}(X - \mathbb{E} X)^2 \quad (5.2.1)$$

**Other Moments** properties:

$$\mathbb{E}(X - \mathbb{E} X + \alpha)^2 = \text{Var } X + \alpha^2 \quad (5.2.2)$$

$$\mathbb{E}(X - \mathbb{E} X + \alpha)^3 = \mathbb{E}(X - \mathbb{E} X)^3 + 3\alpha \text{Var } X + \alpha^3 \quad (5.2.3)$$

where  $\alpha$  is a constant.

**Stochastic dominance** Let  $X$  and  $Y$  be random variables with the corresponding c.d.f-s  $F_X$  and  $F_Y$ .  $X$  1st-order dominates  $Y$  if  $F_X(x) \leq F_Y(x) \quad \forall x$ . Intuitively this means  $Y$  realizations are larger than  $X$  realizations.

**Entropy** For random variable  $X$

$$e(X) = -\mathbb{E} \log f_X = -\int f_X(x) \cdot \log f_X(x) dx \quad (5.2.4)$$

#### 5.2.2 Sufficient statistics

$s(X)$  is a *sufficient statistics* of  $X$  if

$$P(\theta|X) = P(\theta|s(X)) \quad (5.2.5)$$

where  $\theta$  is the parameter for distribution of  $X$ .

### 5.3 Stochastic Boundedness and Convergence in Probability

**Stochastic Boundedness** Random sequence  $\left\{\frac{X_n}{a_n}\right\}$  is stochastically bounded if:  $\forall \epsilon > 0, \exists M > 0, N > 0$ :

$$\Pr(|X_n/a_n| > M) < \epsilon \quad \forall n > N \quad \text{or alternatively:} \quad X_n = O_p(a_n) \quad (5.3.1)$$

**Convergence in Probability** Random sequence  $\{X_n\}$  converges in probability to 0 as  $n \rightarrow \infty$  if

$$\lim_{n \rightarrow \infty} \Pr(|X_n| \geq \epsilon) = 0 \quad \forall \epsilon > 0 \quad \text{or alternatively:} \quad X_n = o_p(1) \quad (5.3.2)$$



**Pointwise Convergence in Probability** Random sequence  $\{Q_n(\theta)\}$  converges in probability pointwise to  $Q(\theta)$  on  $\Theta$  as  $n \rightarrow \infty$  iff

$$\forall \theta \in \Theta, \epsilon > 0, \eta > 0 \quad \exists N : \Pr(|Q_n(\theta) - Q(\theta)| \geq \epsilon) < \eta \quad \forall n > N \quad (5.3.3)$$

or alternatively:  $Q_n(\theta) - Q(\theta) = o_p(1)$ .

**Uniform Convergence in Probability** Random sequence  $\{Q_n(\theta)\}, n = \{1, 2, \dots\}$ ,  $\theta \in \Theta$  converges to  $Q(\theta)$  in probability uniformly iff

$$\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| = o_p(1) \quad (5.3.4)$$

**Stochastic Equicontinuity** Random sequence  $\{Q_n(\theta)\}, n = \{1, 2, \dots\}$ ,  $\theta \in \Theta$  is stochastically equicontinuous if  $\forall \epsilon, \eta > 0$  there exists random  $\Delta_n(\epsilon, \eta)$  and a constant  $n_0(\epsilon, \eta)$  such that

$$\Pr(|\Delta_n(\epsilon, \eta)| > \epsilon) < \eta \quad \forall n > n_0(\epsilon, \eta), \quad (5.3.5)$$

and for each  $\theta$  there exists an open set  $\mathcal{N}(\theta, \epsilon, \eta)$  containing  $\theta$  where

$$\sup_{\tilde{\theta} \in \mathcal{N}(\theta, \epsilon, \eta)} |\hat{Q}_n(\tilde{\theta}) - \hat{Q}_n(\theta)| < \Delta_n(\epsilon, \eta) \quad \forall n > n_0(\epsilon, \eta) \quad (5.3.6)$$

See [Newey \(1991\)](#)

## 5.4 Distributions: General Concepts

### 5.4.1 Sõltumatud juhuslikud muutujad

$X$  ja  $Y$  on sõltumatud  $\Leftrightarrow f(x, y) = f_X(x)f_Y(y) \Leftrightarrow F(x, y) = F_X(x)F_Y(y)$ .

### 5.4.2 Expectations

Let support of random variable  $X$  be  $[a, b]$ . Expectation of  $X$

$$\mathbb{E} X = \int_a^b x \, dF_X(x) \quad (5.4.1)$$

$$= a + \int_a^b \bar{F}_X(x) \, dx. \quad (5.4.2)$$

### Law of iterated expectations

$$\mathbb{E}_X [\mathbb{E}[Y|X]] = \mathbb{E}[Y] \quad (5.4.3)$$

### 5.4.3 Central and Non-Central Moments

Definition:  $n$ -th non-central moment is  $\mu_n = \mathbb{E} X^n$ , and the corresponding central moment is  $m_n = \mathbb{E}(X - \mathbb{E} X)^n$ .

Relationships:

$$\mu_1 \equiv \mu \quad (5.4.4)$$

$$\sigma^2 \equiv m_2 = \mu_2 - \mu^2 \quad (5.4.5)$$

$$m_3 = \mu_3 - 3\mu\mu_2 + 2\mu^3 \quad (5.4.6)$$

$$m_4 = \mu_4 - 4\mu_3\mu + 6\mu_2\mu^2 - 3\mu^4 \quad (5.4.7)$$

#### 5.4.4 Characteristic Function

Let  $X$  be a random variable. It's characteristic function:

$$\phi_X(t) = \mathbb{E} e^{itX} \quad (5.4.8)$$

Properties: for independent random variables  $X_1, X_2$

$$\phi_{X_1+X_2}(t) = \phi_{X_1}(t) \cdot \phi_{X_2}(t) \quad (5.4.9)$$

#### 5.4.5 Momendifunktsioon (*moment generating function*)

MGF avaldub ühemõõtmelisel juhul:

$$M_X(s) = \mathbb{E} e^{sX} = \int e^{sx} f(x) dx. \quad (5.4.10)$$

Momendifunktsiooni omadused:

$$M'_X(0) = \mathbb{E} X \quad M''_X(0) = \mathbb{E} X^2 \quad M^{(n)}_X(0) = \mathbb{E} X^n. \quad (5.4.11)$$

Kasulik asi on ka  $\log X$ -i momendifunktsioon:

$$M_{\log X}(s) = \mathbb{E} e^{s \log X} = \mathbb{E} X^s. \quad (5.4.12)$$

$N$ -mõõtmelisel juhul avaldub MGF:

$$\begin{aligned} M(s_1, s_2, \dots, s_N) &= \mathbb{E} e^{\sum s_i X_i} = \\ &= \int \dots \int e^{s_1 x_1} e^{s_2 x_2} \dots e^{s_N x_N} f(x_1, x_2, \dots, x_N) dx_1 dx_2 \dots dx_N. \end{aligned} \quad (5.4.13)$$

#### 5.4.6 Kumulandifunktsioon (*cumulant-generating function*)

KGF avaldub MGF-i kaudu:

$$K_X(s) = \log M(s) \quad \text{või} \quad K(s_1, s_2, \dots, s_N) = \log M(s_1, s_2, \dots, s_N). \quad (5.4.14)$$

KGF-i omadus (ühemõõtmelisel juhul):

$$K'_X(0) = \mathbb{E} X \quad (5.4.15)$$

$$K''_X(0) = \text{Var } X \quad (5.4.16)$$

$$K'''_X(0) = \mathbb{E}(X - \mathbb{E} X)^3 \quad (5.4.17)$$

ja kahemõõtmelisel juhul:

$$\frac{\partial^2 K(0, 0)}{\partial s_1 \partial s_2} = \text{Cov}(X_1, X_2). \quad (5.4.18)$$

### 5.4.7 Probability Generating Function

For discrete, non-negative random variables

$$G(z) = \mathbb{E}(z^X) = \sum_{x=0}^{\infty} p(x)z^x. \quad (5.4.19)$$

Properties:

$$p(k) = \Pr(X = k) = \frac{G^{(k)}(0)}{k!} \quad (5.4.20)$$

### 5.4.8 Distribution Function of a Function of Random Variable

Olgu juhuslikud muutujad  $X$  ja  $Y$  kusjuures  $X = X(Y)$ . Siis

$$F_x(x) = \Pr[X < x] = \Pr[X < x(y)] = F_x[x(y)] = F_y(y) \quad (5.4.21)$$

ja

$$f_y(y) = F'_y(y) = \frac{d}{dy} F_x[x(y)] = \frac{d}{dx} F_x(x) \frac{dx}{dy} = f_x[x(y)] \frac{dx}{dy}. \quad (5.4.22)$$

### 5.4.9 Information and Entropy

Entropy of random variable  $X$  describes how much information we will gain from an observation of  $X$ :

$$H(X) = - \sum_k \Pr(X = k) \log_2 \Pr(X = k) \quad (5.4.23)$$

A draw from uniform distribution gives maximum possible information of all distributions as the prior is the least informative.

### 5.4.10 Kullback-Leibler divergence

A measure of dissimilarity between distributions  $p$  and  $q$ :

$$KL(p||q) = \mathbb{E}_p \left[ \log \frac{p(X)}{q(X)} \right] = \sum_k p_k \log \frac{p_k}{q_k} \quad (5.4.24)$$

where  $\mathbb{E}_p$  means expectation over  $X$  according to distribution  $p$ . The second equality is true if  $X$  is discrete.

Note: it is not distance as  $KL(p||q) \neq KL(q||p)$ .

### 5.4.11 Mutual Information

Mutual Information (MI) for two random variables  $X$  and  $Y$  is the [Kullback-Leibler divergence](#) between  $P(X, Y)$  and  $P(X)P(Y)$ :

$$MI(X, Y) = KL(P_{XY}(X, Y) || P_X(X)P_Y(Y)) = \sum_x \sum_y p(x, y) \log \frac{P_{XY}(X, Y)}{P_X(X)P_Y(Y)}. \quad (5.4.25)$$

Properties:

- If  $X, Y$  are independent,  $MI(X, Y) = 0$ .
- $MI(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$  where  $H(X|Y) = \sum_x p(x)H(Y|X = x)$ .

## 5.5 One-dimensional discrete distributions

### 5.5.1 Bernoulli

Is the simplest binary distribution: event  $E$  happens with probability  $\mu$  and does not happen with  $1 - \mu$ . The random variable

$$X = \mathbb{1}(E) = \begin{cases} 1 & \text{if } E, \\ 0 & \text{if } \bar{E}. \end{cases} \quad (5.5.1)$$

has Bernoulli distribution.

Properties:

$$\mathbb{E} X = \mu \quad (5.5.2)$$

$$\text{Var } X = \mu(1 - \mu) \quad (5.5.3)$$

Exponential family minimal canonical parameters

$$\eta = \log\left(\frac{\mu}{1 - \mu}\right) \quad \mu = \Lambda(\eta) \quad (5.5.4)$$

$$A(\eta) = \log(1 + e^\eta) = \log\left(\frac{1}{1 - \mu}\right) \quad (5.5.5)$$

$$h(x) = 1 \quad (5.5.6)$$

$$T(x) = x \quad (5.5.7)$$

where  $\Lambda(\cdot)$  is the logistic function.

### 5.5.2 Binoomjaotus

On  $N$  ühesuguse sõltumatu Bernoulli jaotusega juhusliku muutuva summa jaotus. Olgu  $X = \sum^N Y_i$  kus  $Y_i$  on Bernoulli jaotusega parameetriga  $p$ . Siis

$$\Pr(X = x) = \binom{N}{x} p^x (1 - p)^{N-x} \quad x \in \{0, \dots, N\}$$

$$\mathbb{E} X = Np$$

$$\mathbb{E}(X - \mathbb{E} X)^2 = Np(1 - p)$$

$$\mathbb{E}(X - \mathbb{E} X)^3 = Np(1 - p)(1 - 2p)$$

$$\mathbb{E}(X - \mathbb{E} X)^4 = Np(1 - p)(1 - 3p + 3p^2)$$

$$\mathbb{E}(X - \mathbb{E} X)^5 = Np(1 - p)(1 - 2p)(1 - 2p + 2p^2)$$

$$\mathbb{E}(X - \mathbb{E} X)^6 = Np(1 - p)(1 - 5p + 10p^2 - 10p^3 + 5p^4)$$

Tõestus: ühe katse korral kirjuta lahti  $(y - p)^n$ , arvesta et  $Ey^n = p$  ( $i \neq 0$ ).  $N$  sõltumatu katse korral momendid liituvad.

### 5.5.3 Diskreetne jaotus

Jaotus kus juhuslikul muutujal võib olla lõplik hulk diskreetseid väärtusi.

### 5.5.4 Geomeetriline jaotus ( $Geo(p)$ )

Geomeetriline jaotus kirjeldab mingi hulga Bernoulli jaotusega suuruste järjest esinemist. Olgu sündmuse tõenäosus  $p$ . Tõenäosus, et järjest toimub  $n$  sündmust ja seejärel sündmuste jada katkeb on:

$$f(n) = (1 - p)p^n. \quad (5.5.8)$$

Jaotuse omadused:

$$\mathbb{E} n = \frac{1 - p}{p} \quad (5.5.9)$$

$$\text{Var } n = \frac{1 - p}{p^2} \quad (5.5.10)$$

### 5.5.5 Multinoomjaotus

Olgu üksikul katsel  $M$  võimalikku tulemust  $A_1 \dots A_M$  vastavate tõenäosustega  $p_1 \dots p_M$ , kusjuures  $\sum^M p_i = 1$ . Olgu  $N$ -katselises seerias  $N_i$  realiseerunud sündmuste  $A_i$  arv. Siis:

$$\mathbb{E} N_i = N p_i \quad (5.5.11)$$

$$\text{Var } N_i = N p_i (1 - p_i) \quad (5.5.12)$$

$$\dots$$
$$\text{Cov}(N_i, N_j) = -N p_i p_j \quad (5.5.13)$$

Tõestus: momentide arvutamisel võib multinoomjaotuse taandada binoomjaotuseks, kovariatsiooni jaoks kirjutada definitsioon lahti, arvesta et  $\mathbb{E} A_i A_j = 0$ .

### 5.5.6 Poissoni jaotus

Poissoni jaotusega on sõltumatute sündmuste arv ajaühikus. Kui ajaühikus toimub keskmiselt  $\lambda$  sündmust, siis tõenäosus, et toimub  $n$  sündmust aja  $t$  jooksul on:

$$f(n) \equiv p_p(n) = \frac{e^{-\lambda t} (\lambda t)^n}{n!} \quad (5.5.14)$$

$$P_p(n) \equiv \sum_{s=1}^n p_p(s) \quad (5.5.15)$$

Jaotuse omadused: jaotus on log-kumer,

$$\mathbb{E} n = \lambda t \quad \text{Var } n = \lambda t \quad (5.5.16)$$

Tõenäosus, et mingi aja  $t$  jooksul ei toimu ühtegi sündmust,  $f(0)$ , on eksponentsiaalse jaotusega  $\mathcal{E}(\lambda)$ .

ML-hinnang: Kui vaatluse  $i$  jooksul, mille kestus on  $t_i$  toimub  $n_i$  sündmust, siis kõigi vaatluste ML hinnang on:

$$\hat{\lambda} = \frac{\sum n_i}{\sum t_i} \quad \text{ja} \quad \text{Var } \hat{\lambda} = \frac{\sum n_i}{(\sum t_i)^2} \quad (5.5.17)$$

Poissoni summa tuletis aja järgi: Olgu

$$\vartheta(t) = \sum_{s=0}^S Q(s) \frac{(\lambda t)^s}{s!} e^{-\lambda t} = \mathbb{E}_s Q(s) \quad (5.5.18)$$

siis

$$\frac{\partial}{\partial t} \vartheta(t) = \lambda \sum_{s=0}^{S-1} [Q(s+1) - Q(s)] p_p(s) - \lambda Q(S) p_p(S) = \quad (5.5.19)$$

$$= \lambda \sum_{s=1}^S Q(s) [p_p(s-1) - p_p(s)] - \lambda Q(0) p_p(0) \quad (5.5.20)$$

$$\frac{\partial}{\partial t} P_p(s) = -\lambda p_p(s) \quad (5.5.21)$$

$$\frac{\partial}{\partial t} p_p(s) = \begin{cases} -\lambda p_p(s) & \text{kui } s = 0 \\ -\lambda p_p(s) + \lambda p_p(s-1) & \text{kui } s > 0 \end{cases} \quad (5.5.22)$$

### 5.5.7 Skellam Distribution $PD(\lambda, \delta)$

Skellam distribution is distribution of difference of two independent Poisson RV-s. Let  $N = X - Y$  where  $X \sim \text{Poisson}(\lambda)$  and  $Y \sim \text{Poisson}(\delta)$ :

$$P(N = n) = e^{-(\lambda+\delta)} \left( \frac{\lambda}{\delta} \right)^{\frac{n}{2}} I_n(2\sqrt{\lambda\delta}) \quad (5.5.23)$$

$$\mathbb{E} N = \lambda - \delta \quad (5.5.24)$$

$$\text{Var } N = \lambda + \delta \quad (5.5.25)$$

## 5.6 1D Continuous Distributions

### 5.6.1 Beta distribution $\mathcal{B}(a, b)$

$$\text{pdf} \quad f(x) = \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} \quad x \in [0, 1] \quad (5.6.1)$$

$$\text{mode} \quad m = \frac{a-1}{a+b-2} \quad (5.6.2)$$

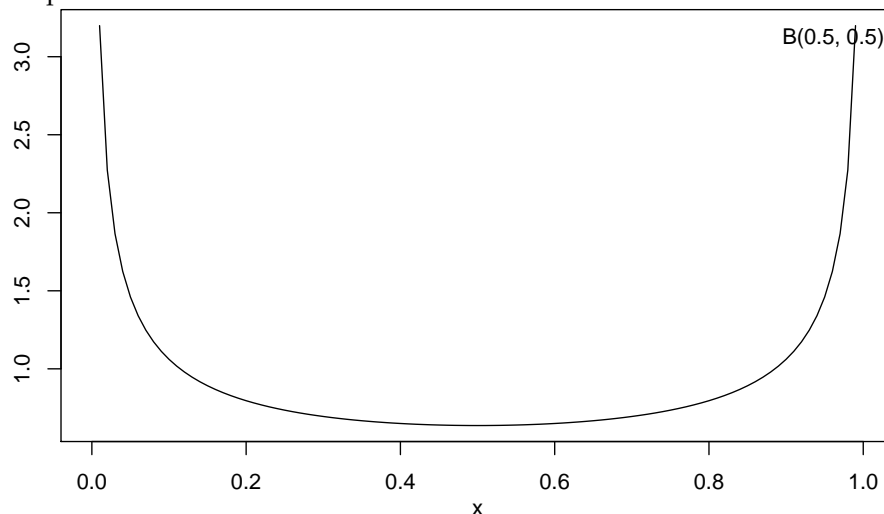
where  $B(a, b)$  is the [beta function \(2.2.10\)](#).

Moments:

Order	Non-Central	Central
1	$\frac{a}{(a+b)}$	0
2		$\frac{ab}{(a+b)^2(a+b+1)}$

Properties:  $B(1, 1)$  is uniform distribution

Examples densities:



### 5.6.2 Eksponentjaotus $\mathcal{E}(\theta)$

Eksponentjaotus kirjeldab konstantse kiirusega hääbuvaid protsesse. Tõenäosustihedus:

$$f(t) = \theta e^{-\theta t}, \quad t \geq 0, \theta > 0 \quad (5.6.3)$$

jaotusfunktsioon:

$$F(t) = 1 - e^{-\theta t}, \quad (5.6.4)$$

momendifunktsioon:

$$M_T(s) = \frac{1}{1 - \frac{s}{\theta}}, \quad s < \theta. \quad (5.6.5)$$

Moments:

Order	Non-Central	Central
1	$\frac{1}{\theta}$	0
2	$\frac{2}{\theta^2}$	$\frac{1}{\theta^2}$
3	$\frac{6}{\theta^3}$	$\frac{2}{\theta^3}$
4	$\frac{24}{\theta^4}$	$\frac{9}{\theta^4}$

$\log T$  on esimest liiki ekstreemväärtuste jaotusega.  $\log T$ -ga seotud suurused on:

$$M_{\log T}(s) = \frac{\Gamma(s+1)}{\theta^s} \quad (5.6.6)$$

$$K_{\log T}(s) = \log \Gamma(s+1) - s \log \theta \quad (5.6.7)$$

$$\mathbb{E} \log T = \psi(1) - \log \theta \quad (5.6.8)$$

$$\text{Var} \log T = \psi'(1), \quad (5.6.9)$$

kus  $\psi$  on digamma funktsioon.

Kui  $z_1 \sim \mathcal{E}(\theta_1)$  ja  $z_2 \sim \mathcal{E}(\theta_2)$  siis

$$\log z_1 - \log z_2 \sim \frac{\theta_1}{\theta_1 + \theta_2 e^{-x}} \sim \Lambda(x), \quad \text{kui } \theta_1 = \theta_2 \quad (5.6.10)$$

Tõestus: arvesta et  $\Pr(z_1/z_2 < \alpha) = \Pr(z_1 < \alpha z_2)$  ja integreeri.

### 5.6.3 F-jaotus $F(n_1, n_2)$

F-jaotus tekib kahe  $\chi^2$  jaotusega suuruse jagamisel. Kui  $w_1 \sim \chi_{n_1}^2$  ja  $w_2 \sim \chi_{n_2}^2$  siis

$$\frac{\frac{w_1}{n_1}}{\frac{w_2}{n_2}} \sim F(n_1, n_2). \quad (5.6.11)$$

Tihedusfunktsioon:

$$f(x) = \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right) \left(1 + \frac{n_1}{n_2} x\right)^{\frac{n_1+n_2}{2}}} \quad (5.6.12)$$

### 5.6.4 Gammajaotus $\mathcal{G}(\alpha, \beta)$

Kasutatakse heterogeensuse kirjeldamiseks kestusmodelites. Tõenäosustihedus:

$$f(x) = \frac{1}{\beta^\alpha} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad x > 0, \quad (5.6.13)$$

where  $\alpha > 0$  is shape and  $\beta > 0$  is scale. Alternatively, one uses parameterization  $(\alpha, \rho = \frac{1}{\beta})$  where  $\rho$  is rate.

Momendifunktsioon ja ootused:

$$M_x(s) = \frac{1}{(\beta s - 1)^\alpha} \quad (5.6.14)$$



$$K_x(s) = -\alpha \log(\beta s - 1) \quad (5.6.15)$$

$$\mathbb{E} x = \beta \alpha \quad (5.6.16)$$

$$\mathbb{E} x^2 = \beta^2 \alpha (\alpha + 1) \quad (5.6.17)$$

$$\text{Var } x = \beta^2 \alpha. \quad (5.6.18)$$

$\log x$  kirjeldav momendifunktsioon ja ootused:

$$M_{\log x}(s) = \frac{\Gamma(s + \alpha)}{\Gamma(\alpha)} \beta^s \quad (5.6.19)$$

$$K_{\log x}(s) = s \log \beta + \log \Gamma(s + \alpha) - \log \Gamma(\alpha) \quad (5.6.20)$$

$$\mathbb{E} \log x = \log \beta + \psi(\alpha) \quad (5.6.21)$$

$$\text{Var } \log x = \psi'(\alpha) \quad (5.6.22)$$

Erijuht on *normaalne gammajaotus*, mille keskväärtus on 1. Sel juhul

$$\alpha = \frac{1}{\beta} \equiv \eta \quad (5.6.23)$$

ja jaotusfunktsioon

$$f_x(x) = \eta^\eta \frac{1}{\Gamma(\eta)} x^{\eta-1} e^{-\eta x}. \quad (5.6.24)$$

Sel juhul:

$$\mathbb{E} x = 1 \quad (5.6.25)$$

$$\mathbb{E} x^2 = 1 + \frac{1}{\eta} \quad (5.6.26)$$

$$\text{Var } x = \frac{1}{\eta} \quad (5.6.27)$$

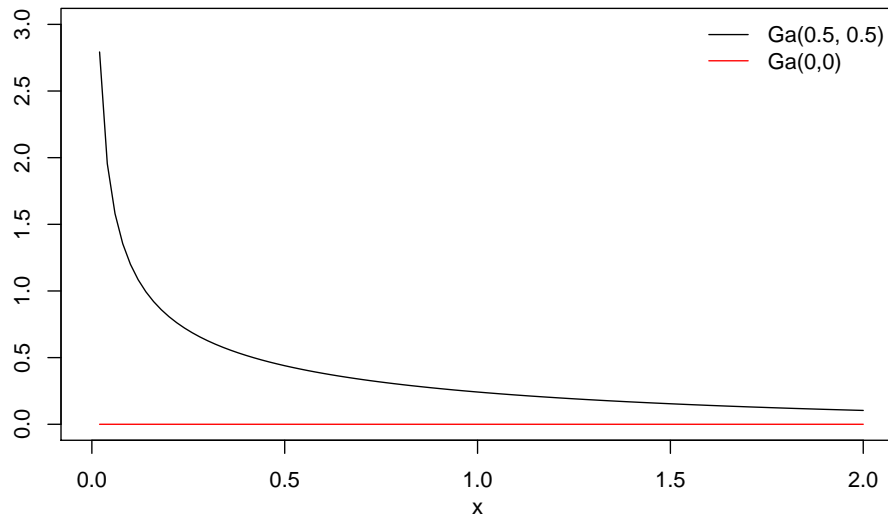
Teine oluline erijuht on  $\chi^2$ -jaotus. Kui  $\alpha = \frac{k}{2}$  ja  $\beta = 2$ , siis  $X$  jaotusfunktsioon on

$$f_x(x) = \frac{1}{2^{\frac{k}{2}}} \frac{1}{\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}. \quad (5.6.28)$$

This is also known as  $\chi^2(k)$  distribution.

Gamma distribution generalized to matrices is [Wishart distribution](#).

Density examples:



### 5.6.5 Hii-ruut jaotus $\chi^2(k)$

$\chi^2(k)$  jaotus tekib kui liita kokku  $k$  normaaljaotusega juhusliku suuruse ruutu. Jaotusfunktsioon:

$$\text{pdf} \quad f_x(x) = \frac{1}{2^{\frac{k}{2}}} \frac{1}{\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}} \quad (5.6.29)$$

$$\text{MGF} \quad M(s) = (1 - 2s)^{-\frac{k}{2}} \quad (5.6.30)$$

### 5.6.6 Inverse gamma distribution

If  $X \sim Ga(a, b)$  is **gamma distributed** then it's inverse  $1/X \sim IG(a, b)$  is inverse-gamma distributed. Properties

$$\text{density} \quad f(x) = \frac{\beta^{-\alpha}}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\frac{\beta}{x}} \quad (5.6.31)$$

### 5.6.7 Laplace distribution

Also double-sided exponential distribution. Density:

$$f_X(x; \mu, b) = \frac{1}{2b} e^{-\frac{|x - \mu|}{b}}. \quad (5.6.32)$$

Properties:

$$\mathbb{E} X = \mu \quad \text{Var } X = 2b^2 \quad (5.6.33)$$

### 5.6.8 Log-normal $LN(\mu, \sigma^2)$

Distribution of RV  $X$  if  $\log X \sim N()$ .

$$\text{cdf} \quad f_x(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{1}{2} \left[ \frac{(\log x - \mu)}{\sigma} \right]^2} \quad (5.6.34)$$

Properties:

$$\mathbb{E} x = e^{\mu + \frac{1}{2}\sigma^2} \quad (5.6.35)$$

$$\text{Var } x = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) \quad (5.6.36)$$

$$\text{median} = e^\mu \quad (5.6.37)$$

### 5.6.9 Log-ühtlane jaotus

Kasutatakse palgajaotuse kirjeldamiseks.

$$\text{pdf} \quad f(x) = \frac{1}{x} \cdot \frac{1}{\log \beta - \log \alpha} \quad \text{where} \quad 0 \leq \alpha \leq x \leq \beta < \infty. \quad (5.6.38)$$

### 5.6.10 Logistic Distribution

$$\text{cdf} \quad \Lambda(x) = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}} \quad (5.6.39)$$

$$\begin{aligned} \text{pdf} \quad f(x) &= \frac{e^x}{(1 + e^x)^2} = \frac{e^{-x}}{(1 + e^{-x})^2} = \\ &= \Lambda(-x)\Lambda(x) = [1 - \Lambda(x)]\Lambda(x) \end{aligned} \quad (5.6.40)$$

$$f'(x) = e^{-x} \frac{e^{-x} - 1}{(e^{-x} + 1)^3} \quad (5.6.41)$$

$$\text{MGF} \quad M(s) = \int \frac{e^{x(s+1)}}{(1 + e^x)^2} dx \quad (5.6.42)$$

Logistic distribution is symmetric around 0, i.e.  $\Lambda(x) = 1 - \Lambda(-x)$  and  $f(x) = f(-x)$ .

### 5.6.11 Normal Distribution $N(\mu, \sigma^2)$

Sum of many independent random disturbances tends to be normally distributed (Central Limit Theorem.)

$$F(x; \mu, \sigma) \equiv \Phi\left(\frac{x - \mu}{\sigma}\right) \quad \text{cannot be expressed analytically} \quad (5.6.43)$$

$$f(x; \mu, \sigma) \equiv \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}} \quad (5.6.44)$$

Moments (from wikipedia):

Order	Non-central moment	Central moment
1	$\mu$	0
2	$\mu^2 + \sigma^2$	$\sigma^2$
3	$\mu^3 + 3\mu\sigma^2$	0
4	$\mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$	$3\sigma^4$
5	$\mu^5 + 10\mu^3\sigma^2 + 15\mu\sigma^4$	0
6	$\mu^6 + 15\mu^4\sigma^2 + 45\mu^2\sigma^4 + 15\sigma^6$	$15\sigma^6$
7	$\mu^7 + 21\mu^5\sigma^2 + 105\mu^3\sigma^4 + 105\mu\sigma^6$	0
8	$\mu^8 + 28\mu^6\sigma^2 + 210\mu^4\sigma^4 + 420\mu^2\sigma^6 + 105\sigma^8$	$105\sigma^8$

Characteristic function:

$$\phi_X(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2} \quad (5.6.45)$$

Moment generating function

$$M_X(s) = e^{\mu s + \frac{1}{2}\sigma^2 s^2} \quad (5.6.46)$$

Properties: if  $X_i \sim N(\mu_i, \sigma_i^2)$  are independent normals

$$\sum_i X_i \sim N\left(\sum_i \mu_i, \sum_i \sigma_i^2\right). \quad (5.6.47)$$

**Conditional Expectations** If  $X \sim N(\mu, \sigma)$ ,

$$\mathbb{E}[X|X < a] = \mu - \sigma \frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{\Phi\left(\frac{a-\mu}{\sigma}\right)} = \mu - \sigma \lambda\left(\frac{a-\mu}{\sigma}\right) \quad (5.6.48)$$

$$\mathbb{E}[X|X > a] = \mu + \sigma \frac{\phi\left(\frac{a-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{a-\mu}{\sigma}\right)} = \mu + \sigma \lambda\left(\frac{\mu-a}{\sigma}\right) \quad (5.6.49)$$

$$\mathbb{E}[X|X \in [a, b]] = \mu - \sigma \frac{\phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \quad (5.6.50)$$

If  $X \sim N(0, \sigma)$ ,

$$\mathbb{E}[X|X < a] = -\sigma \lambda\left(\frac{a}{\sigma}\right) \quad (5.6.51)$$

$$\mathbb{E}[X|X > a] = \sigma \lambda\left(-\frac{a}{\sigma}\right) \quad (5.6.52)$$

$$\mathbb{E}[X^2|X < a] = \sigma^2 - \sigma a \lambda\left(\frac{a}{\sigma}\right) \quad (5.6.53)$$

$$\mathbb{E}[X^2|X > a] = \sigma^2 + \sigma a \lambda\left(-\frac{a}{\sigma}\right) \quad (5.6.54)$$

$$\mathbb{E}[X^2|X > -a \wedge X < a] = \sigma^2 - 2 \frac{\sigma a \phi\left(\frac{a}{\sigma}\right)}{1 - 2\Phi\left(\frac{a}{\sigma}\right)} \quad (5.6.55)$$

$$\mathbb{E}[X^2|X < -a \vee X > a] = \mathbb{E}[X^2|X > a] \quad (5.6.56)$$

$$\text{Var}[X|X < a] = \sigma^2 \left[1 - \frac{a}{\sigma} \lambda\left(\frac{a}{\sigma}\right) - \lambda^2\left(\frac{a}{\sigma}\right)\right] \quad (5.6.57)$$

$$\text{Var}[X|X > a] = \sigma^2 \left[1 + \frac{a}{\sigma} \lambda\left(-\frac{a}{\sigma}\right) - \lambda^2\left(-\frac{a}{\sigma}\right)\right] \quad (5.6.58)$$

$$(5.6.59)$$

Let  $X \sim N(\mu_X, \sigma_X^2)$  and  $Y \sim N(\mu_Y, \sigma_Y^2)$ ,  $X \perp Y$ . Now

$$\mathbb{E}[X|X < Y] = \mu_X - \frac{\sigma_X^2}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \lambda\left(-\frac{\mu_X - \mu_Y}{\sqrt{\sigma_X^2 + \sigma_Y^2}}\right) \quad (5.6.60)$$

$$\mathbb{E}[X|X > Y] = \mu_X + \frac{\sigma_X^2}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \lambda \left( \frac{\mu_X - \mu_Y}{\sqrt{\sigma_X^2 + \sigma_Y^2}} \right). \quad (5.6.61)$$

Proof: write  $\mathbb{E}[X|X < Y] = \mathbb{E}[X|Z < 0]$  where  $Z = X - Y$ . Now follows from (5.7.13)

### 5.6.12 Pareto Distribution

Describes the upper part of many highly unequal distributions.

$$F_X(x) = 1 - \left( \frac{x_0}{x} \right)^\alpha, \quad x \geq x_0 > 0; \quad \alpha > 0 \quad (5.6.62)$$

$$f_X(x) = \alpha x_0^\alpha x^{-\alpha-1} \quad (5.6.63)$$

$$\mathbb{E} X = \frac{\alpha}{\alpha - 1} x_0, \quad \alpha > 1 \quad (5.6.64)$$

Properties:

- power law:  $\log f_X(x) = \log(\alpha x_0^\alpha) - (\alpha + 1) \log x$  is linear on log-log scale.
- it is *scale-free*: there is no features in the right tail, wherever you look, you have most observations that are smaller, but you also have observations that are way larger.

### 5.6.13 Pööratud normaaljaotus

Tihedusfunktsioon:

$$f(t) = \frac{1}{t^{\frac{3}{2}}} \phi \left( \frac{\mu t - 1}{\sigma \sqrt{t}} \right) \quad (5.6.65)$$

ja jaotusfunktsioon:

$$F(t) = \Phi \left( \frac{\mu t - 1}{\sigma \sqrt{t}} \right) - e^{2\frac{\mu}{\sigma^2}} \Phi \left( -\frac{\mu t + 1}{\sigma \sqrt{t}} \right). \quad (5.6.66)$$

Kõik momendid on olemas kui  $\mu > 0$ :

$$\mathbb{E} T = \frac{1}{\mu} \quad (5.6.67)$$

$$\text{Var } T = \frac{\sigma^2}{\mu^3}. \quad (5.6.68)$$

Kui  $\mu = 0$ , on jaotus korralik, positiivsed momendid aga puuduvad.

### 5.6.14 t-Distribution

Used for *t-test*. For *n degrees of freedom*:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad (5.6.69)$$

$$\mathbb{E} X = 0 \quad (5.6.70)$$

$$\text{Var } X = \frac{n}{n-2} \quad (5.6.71)$$

$$\text{skewness } g_1 = 0 \quad (5.6.72)$$

$$\text{curtosis } g_2 = \frac{3n-6}{n-4} \quad (n > 4) \quad (5.6.73)$$

#### 5.6.15 Triangular Distribution

$$F(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases} \quad (5.6.74)$$

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases} \quad (5.6.75)$$

$$\mathbb{E} X = \frac{2}{3} \quad (5.6.76)$$

$$\mathbb{E} X^2 = \frac{1}{2} \quad (5.6.77)$$

$$\text{Var } X = \frac{1}{18} \quad (5.6.78)$$

#### 5.6.16 Type-1 Extreme Value Distribution $EV_1$

(Also Gumbel distribution.) Describes  $\log T$  when  $T \sim \mathcal{E}(1)$ .

$$f(x) = e^{-x} e^{-e^{-x}} \quad (5.6.79)$$

$$F(x) = e^{-e^{-x}}. \quad (5.6.80)$$

$$\mathbb{E} x = -c \approx 0,5772 \quad (5.6.81)$$

#### 5.6.17 Type-2 Extreme Value Distribution $EV_2$

(Also Fréchet distribution or inverse Weibull distribution.)

$$f(x) = \alpha x^{-1-\alpha} e^{-x^{-\alpha}} \quad (5.6.82)$$

$$F(x) = e^{-x^{-\alpha}} \quad (5.6.83)$$

#### 5.6.18 Uniform Distribution $Unif(a, b)$

$$\text{pdf} \quad f(x) = \frac{1}{b-a}, \quad a \leq x \leq b \quad (5.6.84)$$

$$\text{MGF} \quad M(s) = \frac{e^{tb} - e^{ta}}{t(b-a)} \quad (5.6.85)$$

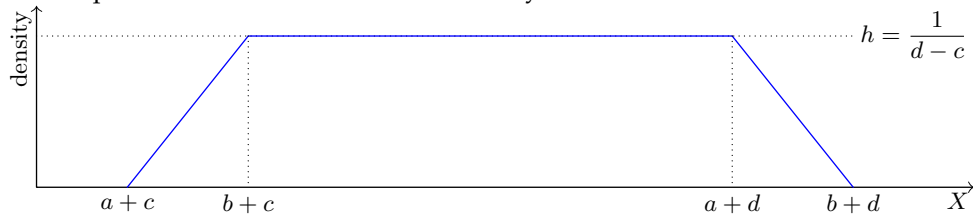
Moments:

Order	Non-Central	Central
1	$\frac{1}{2}(a+b)$	0
2	$\frac{1}{3}(a^2+ab+b^2)$	$\frac{1}{12}(b-a)^2$
3	$\frac{1}{4} \frac{b^4-a^4}{b-a} = \frac{1}{4}(b^2+a^2)(b+a)$	
4	$\frac{1}{5} \frac{b^5-a^5}{b-a}$	$\frac{1}{80}(b-a)^4$

Sum of RV-s with uniform distribution: Let

$$X \sim Unif(a, b) \quad Y \sim Unif(c, d) \quad (5.6.86)$$

be independent and  $d - c \geq b - a$ . The density of  $Z = X + Y$  is



### 5.6.19 Weibulli jaotus

Weibulli jaotus on eksponentjaotuse üldistus, kasutatakse ajas ühtlaselt kahaneva hasardi kirjeldamiseks. Omadused:

$$F(t) = 1 - e^{-(\lambda t)^\alpha} \quad (5.6.87)$$

$$f(t) = \alpha \lambda^\alpha t^{\alpha-1} e^{-(\lambda t)^\alpha} \quad (5.6.88)$$

$$\theta(t) = \alpha \lambda^\alpha t^{\alpha-1} \quad (5.6.89)$$

where  $\lambda$  is scale- and  $\alpha$  is the shape parameter.

## 5.7 Multivariate Continuous Distributions

### 5.7.1 Dirichlet distribution

Multivariate generalization of [beta distribution](#).

$$\text{pdf} \quad f(\mathbf{x}; \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^K x_k^{\alpha_k-1} \mathbb{1}(\mathbf{x} \in S_K) \quad (5.7.1)$$

where  $S_K$  is the probability simplex:

$$S_K = \{\mathbf{x} : 0 \leq x_k \leq 1, \sum_{k=1}^K x_k = 1\} \quad (5.7.2)$$

and  $B(\boldsymbol{\alpha})$  is  $K$ -variable generalization of the beta function:

$$B(\boldsymbol{\alpha}) = \frac{\prod_{k=1}^K \Gamma(\alpha_k)}{\Gamma(\alpha_0)} \quad \alpha_0 = \sum_{k=1}^K \alpha_k \quad (5.7.3)$$

Mode:

$$m_k = \frac{\alpha_k - 1}{\alpha_0 - K} \quad (5.7.4)$$

Moments:

Order	Non-Central	Central
1	$\frac{\alpha_k}{\alpha_0}$	0
2		$\frac{\alpha_k(\alpha_0 - \alpha_k)}{\alpha_0^2(\alpha_0 + 1)}$

### 5.7.2 Multivariate Normal $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

N-mõõtmelise normaajaotuse jaotusfunktsioon: Olgu

$$\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad (5.7.5)$$

kus  $\boldsymbol{\mu}$  on keskväärtus ja  $\boldsymbol{\Sigma}$  dispersioonimaatriks. Siis:

$$f_{\mathbf{X}}(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}. \quad (5.7.6)$$

**2D Conditional Distributions** Let  $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$ .

The exponent in the distribution function can be expressed as

$$-\frac{1}{2} \left( \frac{\sigma_2 x_1^2 - 2\sigma_{12} x_1 x_2 + \sigma_1^2 x_2^2}{\sigma_1^2 \sigma_2^2 - \sigma_{12}^2} \right) = -\frac{1}{2} \left[ \frac{\left( x_1 - \frac{\sigma_{12}}{\sigma_2^2} x_2 \right)^2}{\sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}} + \frac{x_2^2}{\sigma_2^2} \right]. \quad (5.7.7)$$



Accordingly, based on Bayesian law the probability density of 2D normal  $f_{X_1, X_2}(x_1, x_2) = f_{X_1|X_2}(x_1, x_2)f_{X_2}(x_2) = f_{X_2|X_1}(x_1, x_2)f_{X_1}(x_1)$  where all the conditional and marginal distribution functions are normal:

$$(X_1|X_2 = x_2) \sim N\left(\frac{\sigma_{12}}{\sigma_2^2}x_2, \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}\right) \quad X_2 \sim N(0, \sigma_2^2) \quad (5.7.8)$$

$$(X_2|X_1 = x_1) \sim N\left(\frac{\sigma_{12}}{\sigma_1^2}x_1, \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}\right) \quad X_1 \sim N(0, \sigma_1^2) \quad (5.7.9)$$

Distribution for  $(X_1|X_2 < a)$ :

$$f_{X_1|X_2 < a}(x_1) = \frac{1}{\sigma_1} \frac{\phi\left(\frac{x_1}{\sigma_1}\right)}{\Phi\left(\frac{a}{\sigma_2}\right)} \Phi\left(\frac{a - \frac{\sigma_{12}}{\sigma_1^2}x_1}{\sqrt{\sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}}}\right) \quad (5.7.10)$$

Distribution for  $(X_1|X_2 > a)$ :

$$f_{X_1|X_2 > a}(x_1) = \frac{1}{\sigma_1} \frac{\phi\left(\frac{x_1}{\sigma_1}\right)}{\Phi\left(-\frac{a}{\sigma_2}\right)} \Phi\left(-\frac{a - \frac{\sigma_{12}}{\sigma_1^2}x_1}{\sqrt{\sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}}}\right) \quad (5.7.11)$$

**Conditional Expectations** Let  $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N(\mu, \Sigma)$ , where  $\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$  and  $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ :

$$(X_1|X_2 = x_2) \sim N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(x_2 - \mu_2), \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}\right) \quad (5.7.12)$$

(follows from 5.7.7) and

$$\mathbb{E}[X_1|X_2 < a] = \mu_1 - \frac{\sigma_{12}}{\sigma_2} \lambda\left(\frac{a - \mu_2}{\sigma_2}\right) \quad (5.7.13)$$

$$\mathbb{E}[X_1|X_2 > a] = \mu_1 + \frac{\sigma_{12}}{\sigma_2} \lambda\left(\frac{\mu_2 - a}{\sigma_2}\right) \quad (5.7.14)$$

$$\mathbb{E}[X_1^2|X_2 < a] = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^3} a \lambda\left(\frac{a}{\sigma_2}\right) \quad (5.7.15)$$

$$\mathbb{E}[X_1^2|X_2 > a] = \sigma_1^2 + \frac{\sigma_{12}^2}{\sigma_2^3} a \lambda\left(-\frac{a}{\sigma_2}\right) \quad (5.7.16)$$

$$\mathbb{E}[X_1^2|X_2 \in \mathcal{A}] = \frac{\sigma_{12}^2}{\sigma_2^4} \mathbb{E}[X_2^2|X_2 \in \mathcal{A}] + \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} \quad (5.7.17)$$

$$\text{Var}[X_1|X_2 < a] = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^3} a \lambda\left(\frac{a}{\sigma_2}\right) - \frac{\sigma_{12}^2}{\sigma_2^2} \lambda^2\left(\frac{a}{\sigma_2}\right) \quad (5.7.18)$$

$$\text{Var}[X_1|X_2 > a] = \sigma_1^2 + \frac{\sigma_{12}^2}{\sigma_2^3} a \lambda\left(-\frac{a}{\sigma_2}\right) - \frac{\sigma_{12}^2}{\sigma_2^2} \lambda^2\left(-\frac{a}{\sigma_2}\right) \quad (5.7.19)$$

Proof: write (5.7.12)  $\Rightarrow X_1 = \mu_1 + \frac{\sigma_{12}}{\sigma_2}(X_2 - \mu_2) + E$ , where  $E$  and  $X_2$  are independent (property of normal distribution). Find  $X_1|X_2 \in \mathcal{A} = \frac{\sigma_{12}}{\sigma_2^2} \mathbb{E}[X_2|X_2 \in \mathcal{A}] + E$ .

### Multiplying normals

$$\frac{1}{\sigma_1} \phi\left(\frac{x - ay}{\sigma_1}\right) \frac{1}{\sigma_2} \phi\left(\frac{y - b}{\sigma_2}\right) = \frac{1}{\sigma_x} \phi\left(\frac{x - ab}{\sigma_x}\right) \frac{1}{\sigma_y} \phi\left(\frac{y - \mu_y}{\sigma_y}\right), \quad (5.7.20)$$

where

$$\sigma_x^2 = \sigma_1^2 + \sigma_2^2 a^2 \quad \sigma_y = \frac{\sigma_1 \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2 a^2}}$$

$$\mu_y = \frac{\sigma_1^2 b + \sigma_2^2 a x}{\sigma_1^2 + \sigma_2^2 a^2}$$

The same in multi-dimensional case:

$$\begin{aligned} \frac{1}{\sigma_1} \phi\left(\frac{x_1 - y}{\sigma_1}\right) \frac{1}{\sigma_1} \phi\left(\frac{x_2 - y}{\sigma_1}\right) \dots \frac{1}{\sigma_1} \phi\left(\frac{x_n - y}{\sigma_1}\right) \frac{1}{\sigma_2} \phi\left(\frac{y}{\sigma_2}\right) &= \\ &= \prod_{i=1}^n \frac{1}{\sigma_1} \phi\left(\frac{x_i - y}{\sigma_1}\right) \frac{1}{\sigma_2} \phi\left(\frac{y}{\sigma_2}\right) = \\ &= \frac{1}{\sigma_x} \phi\left(\frac{x_1}{\sigma_x}\right) \frac{1}{\sigma_x} \phi\left(\frac{x_2}{\sigma_x}\right) \dots \frac{1}{\sigma_x} \phi\left(\frac{x_n}{\sigma_x}\right) \frac{1}{\sigma_y} \phi\left(\frac{y - \mu_y}{\sigma_y}\right) = \\ &= \prod_{i=1}^n \frac{1}{\sigma_x} \phi\left(\frac{x_i}{\sigma_x}\right) \frac{1}{\sigma_y} \phi\left(\frac{y - \mu_y}{\sigma_y}\right) \end{aligned} \quad (5.7.21)$$

where

$$\sigma_x^2 = \sigma_1^2 \frac{\sigma_1^2 + n\sigma_2^2}{\sigma_1^2 + (n-1)\sigma_2^2} \quad \sigma_y = \frac{\sigma_1 \sigma_2}{\sqrt{\sigma_1^2 + n\sigma_2^2}}$$

$$\mu_y = \frac{\sigma_2^2 \sum_{i=1}^n x_i}{\sigma_1^2 + n\sigma_2^2}$$

### 5.7.3 Wishart distribution

Wishart distribution is a generalization of [gamma distribution](#) to positive definite matrices.

Probability density:

$$f(\mathbf{X}) = \frac{1}{Z_{Wi}} |\mathbf{X}|^{\frac{1}{2}(\nu-D-1)} e^{-\frac{1}{2} \text{Tr}(\mathbf{X}\mathbf{S}^{-1})} \quad (5.7.22)$$

where  $\nu$  is degrees of freedom,  $\mathbf{S}$  is scale matrix, and

$$Z_{Wi} = 2^{\frac{1}{2}\nu D} \Gamma_D\left(\frac{\nu}{2}\right) |\mathbf{S}|^{\frac{\nu}{2}}. \quad (5.7.23)$$

here  $\Gamma_d(\cdot)$  is [multivariate gamma](#) function.

Properties:

$$\mathbb{E} X = \nu S \quad (5.7.24)$$

$$\text{mode} = (\nu - D - 1)S \quad \text{for } \nu > D + 1 \quad (5.7.25)$$

## 5.8 Distribution Families

### 5.8.1 Exponential Family

Exponential family is a distribution whose density can be written as

$$p(x|\eta) = h(x)e^{\eta^T T(x) - A(\eta)} = \frac{h(x)}{Z(\eta)} e^{\eta^T T(x)}, \quad (5.8.1)$$

where  $\eta$  is the canonical parameter,  $T(x)$  sufficient statistic, and  $A(\eta) = \log Z(\eta)$  is the cumulant function. Instead of the canonical parameter, one can use another parameter  $\theta$ :  $\eta = \eta(\theta)$ :

$$p(x|\theta) = h(x)e^{\eta(\theta)^T T(x) - A(\eta(\theta))}. \quad (5.8.2)$$

### 5.8.2 Stabiilne pere

Mittenegatiivsesse stabiilsesse perre kuuluvad jaotused, mille momendifunktsioon on

$$M_x(s) = e^{-s^\alpha}, \quad 0 < \alpha \leq 1. \quad (5.8.3)$$

Stabiilse pere omadused:

1. kui juhusliku muutuja  $X_i$  jaotusfunktsioon on  $G_\alpha$  mis kuulub stabiilsesse perre, siis juhusliku muutuja

$$Y = n^{-\frac{1}{\alpha}} \sum_{i=1}^N X_i \quad (5.8.4)$$

jaotusfunktsioon on kah  $G_\alpha$ .

2. Momendifunktsiooni tuletis

$$M'_x(s) = -\alpha s^{\alpha-1} M(s) \quad (5.8.5)$$

## 6 Estimators

### 6.1 M-Estimators

#### 6.1.1 Variance

Let an estimator solve

$$H = \sum_i h_i(\hat{\theta}) = 0. \quad (6.1.1)$$

From Taylor approximation

$$\sum_i h_i(\hat{\theta}) = \sum_i h_i(\theta_0) + \frac{\partial}{\partial \theta} \sum_i h_i(\theta) \Big|_{\theta_0} (\hat{\theta} - \theta_0) = 0 \quad (6.1.2)$$

from where

$$\hat{\theta} - \theta_0 = - \left( \frac{\partial}{\partial \theta} \sum_i h_i(\theta) \Big|_{\theta_0} \right)^{-1} \sum_i h_i(\theta_0). \quad (6.1.3)$$

The estimate for variance is

$$\text{Var } \hat{\theta} = \hat{A}^{-1} \widehat{\text{Var } H} \hat{A}^{-1} \quad (6.1.4)$$

where

$$\hat{A} = \frac{\partial}{\partial \theta} \sum_i h_i(\theta) \Big|_{\theta_0} \quad (6.1.5)$$

and  $\widehat{\text{Var } H}$  is an estimator for  $\text{Var } H$ .

### 6.2 Maximum likelihood

#### 6.2.1 Definition

Let the random variables  $X_1, X_2, \dots, X_n$  be *i.i.d.* distributed according to a distribution function  $F(\cdot|\vartheta)$  and corresponding density function  $f(\cdot|\vartheta)$ . Let  $f(\cdot|\vartheta)$  be specified fully parametrically with a finite unknown parameter vector  $\vartheta$ . The *log-likelihood* function of the observed values  $x_1, x_2, \dots, x_n$  is:

$$\ell(\vartheta|x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n \log f(x_i|\vartheta). \quad (6.2.1)$$

The *score* is defined as

$$g(\vartheta|x_1, x_2, \dots, x_n) = \frac{\partial}{\partial \theta} \ell(\vartheta|x_1, x_2, \dots, x_n) \quad (6.2.2)$$

The *maximum likelihood* estimator of  $\vartheta$  is the value of  $\vartheta$  which maximises the log-likelihood function:

$$\hat{\vartheta} = \arg \max_{\vartheta} \ell(\vartheta|x_1, x_2, \dots, x_n). \quad (6.2.3)$$

### 6.2.2 Information matrix

Information matrix is defined as

$$I(\boldsymbol{\vartheta}) \equiv -\mathbb{E} \left[ \frac{\partial^2 \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}'} \right] = \mathbb{E} \left[ \frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}} \frac{\partial \ell(\boldsymbol{\vartheta})}{\partial \boldsymbol{\vartheta}'} \right]. \quad (6.2.4)$$

(information equality.)

### 6.2.3 Relationship to Kullback-Leibler divergence

ML estimator can be written as

$$\hat{\boldsymbol{\vartheta}} = \arg \max_{\boldsymbol{\vartheta}} \int \log f(x|\boldsymbol{\vartheta}) dF_n(x), \quad (6.2.5)$$

where  $F_n(x)$  is the empirical distribution function:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(X_i \leq x). \quad (6.2.6)$$

Further, we may write the estimator as

$$\hat{\boldsymbol{\vartheta}} = \arg \min_{\boldsymbol{\vartheta}} \left[ \int \log f(x|\boldsymbol{\vartheta}_0) dF_n(x) - \int \log f(x|\boldsymbol{\vartheta}) dF_n(x) \right], \quad (6.2.7)$$

where  $f(\cdot|\boldsymbol{\vartheta}_0)$  is the true density function of  $X$  that does not depend on  $\boldsymbol{\vartheta}$ . Hence

$$\hat{\boldsymbol{\vartheta}} = \arg \min_{\boldsymbol{\vartheta}} \int \log \frac{f(x|\boldsymbol{\vartheta}_0)}{f(x|\boldsymbol{\vartheta})} dF_n(x) = \arg \min_{\boldsymbol{\vartheta}} KL(f|\boldsymbol{\vartheta}_0||f|\boldsymbol{\vartheta}), \quad (6.2.8)$$

the [Kullback-Leibler divergence](#) of distributions  $f|\boldsymbol{\vartheta}_0$  and  $f|\boldsymbol{\vartheta}$ .

## 6.3 Generalized Method of Moments

The asymptotic variance of the estimator is

$$V_{GMM} = \frac{1}{n} [\boldsymbol{\Gamma}' \mathbf{W} \boldsymbol{\Gamma}]^{-1} \quad (6.3.1)$$

where  $\mathbf{W}$  is the weighting matrix and

$$\boldsymbol{\Gamma} = \frac{\partial \bar{m}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \quad (6.3.2)$$

### Optimal Weighting Matrix

$$\mathbf{W}^* = \frac{1}{n} \left\{ \text{Asy. Var} \left[ \frac{1}{n} \sum_i m_i \right] \right\}^{-1} \quad (6.3.3)$$

### 6.3.1 Optimal Weighting Matrices and Asymptotic Variances

Assume we have i.i.d sample of random values  $X_i$ . Let  $\mathbb{E} X = \mu$ ,  $\text{Var} X = \sigma^2$ ,  $\mathbb{E} X^2 = \mu_2 = \mu^2 + \sigma^2$ ,  $\mathbb{E} X^4 = \mu_4$ .

For the moment condition

$$\mathbb{E} X - \mu = 0 \quad (6.3.4)$$

the optimal weighting matrix:

$$W = \frac{1}{n} \left( \frac{\mu_2 - \mu^2}{n} \right)^{-1} = \frac{1}{\sigma^2} \quad \text{and} \quad \text{Var} \hat{\mu} = \frac{1}{n} \sigma^2 \quad (6.3.5)$$

For the moment condition

$$\mathbb{E} X^2 - \mu_2 = 0 \quad (6.3.6)$$

the optimal weighting matrix:

$$W = \frac{1}{n} \left( \frac{\mu_4 - \mu_2^2}{n} \right)^{-1} = \frac{1}{\mu_4 - \mu_2^2} \quad \text{and} \quad \text{Var} \hat{\mu} = \frac{1}{n} \frac{\mu_4 - \mu_2^2}{4\mu^2} \quad (6.3.7)$$

For the moment condition

$$\begin{pmatrix} \mathbb{E} X - \mu \\ \mathbb{E} X^2 - \mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{or} \quad \bar{m} = \begin{pmatrix} \frac{1}{n} \sum_i X_i - \mu \\ \frac{1}{n} \sum_i X_i^2 - \mu_2 \end{pmatrix}. \quad (6.3.8)$$

Matrix

$$\Gamma = \frac{\partial \bar{m}((\mu, \sigma^2)')}{\partial (\mu, \sigma^2)} = \begin{pmatrix} -1 & 0 \\ -2\mu & -1 \end{pmatrix} \quad (6.3.9)$$

we have the optimal weighting matrix:

$$W^{-1} = \begin{pmatrix} \mu_2 - \mu^2 & \mu_3 - \mu\mu_2 \\ \mu_3 - \mu\mu_2 & \mu_4 - \mu_2^2 \end{pmatrix} = \begin{pmatrix} \sigma^2 & \mu_3 - \mu\mu_2 \\ \mu_3 - \mu\mu_2 & \mu_4 - \mu_2^2 \end{pmatrix} \quad (6.3.10)$$

and the variance

$$V = \frac{1}{n} \begin{pmatrix} \sigma^2 & -2\mu\sigma^2 - \mu_2\mu + \mu_3 \\ -2\mu\sigma^2 - \mu_2\mu + \mu_3 & -4\mu^2\sigma^2 - 4\mu\mu_3 + 2\mu^2\mu_2 + \mu_4 - \mu_2^2 \end{pmatrix} \quad (6.3.11)$$

If  $\mu_3 = \mu = 0$ , the variance is

$$V = \frac{1}{n} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \mu_4 - \sigma^2 \end{pmatrix} \quad (6.3.12)$$

## 6.4 Entropy Distance

### 6.4.1 Entropy Distance

[Lin \(1991\)](#) defines *entropy distance*:

$$K(p_1, p_2) = \sum_x p_1(x) \log_2 \frac{p_1(x)}{\frac{1}{2}[p_1(x) + p_2(x)]} = 1 + \frac{1}{\log 2} \sum_x p_1(x) \log \frac{p_1(x)}{p_1(x) + p_2(x)} \quad (6.4.1)$$

Properties:  $K(p_1, p_2) = 0$  if and only if  $p_1 \equiv p_2$ . Otherwise,  $K > 0$ .

## 7 Stochastic Processes

### 7.1 Autoregressive (AR) Processes

**AR(1) process** Juhuslik muutuja  $U$  järgib AR(1) protsessi kui  $U$  käesoleva perioodi realisatsioon on seotud eelmise perioodi omaga

$$u_t = \varrho u_{t-1} + \varepsilon_t \quad (7.1.1)$$

ning  $\varepsilon_t$  väärtused eri ajaperioodidel on sõltumatud. Et protsess oleks stabiilne peab  $\varrho$  väärtus jääma vahemikku  $(-1, 1)$ .

**AR(2) protsess** Juhuslik muutuja  $U$  järgib AR(2) protsessi kui  $U$  käesoleva perioodi realisatsioon on seotud kahe eelmise perioodi omaga

$$u_t = \varrho_1 u_{t-1} + \varrho_2 u_{t-2} + \varepsilon_t \quad (7.1.2)$$

ning  $\varepsilon_t$  väärtused eri ajaperioodidel on sõltumatud.

### 7.2 Hulkumine

Definitsioon: hulkumine (*random walk*) on statistiline protsess

$$z_{t+1} = z_t + \varepsilon_{t+1}. \quad (7.2.1)$$

#### 7.2.1 Hulkumine vastu barjääri

Olgu  $z_0 = 0$  ja  $\varepsilon \sim N(0, 1)$  *i.i.d.* protsess. Siis  $z_2$  jaotus tingimusel et  $z_1$  es ülela barjääri  $\alpha$  on

$$f(z_2|z_1 < \alpha) = \frac{1}{\sqrt{2}} \frac{\Phi\left(\sqrt{2}\alpha - \frac{1}{\sqrt{2}}z_2\right)}{\Phi(\alpha)} \phi\left(\frac{z_2}{\sqrt{2}}\right) \quad (7.2.2)$$

Tõestus: kirjuta  $\phi(x)$  lahti ja integreeri.

## 8 Statistilised mudelid

### 8.1 Tobit-2 model

Definition:

$$y_{1i}^* = z_i' \gamma + u_{1i} \quad (8.1.1)$$

$$y_{2i}^* = x_i' \beta + u_{2i} \quad (8.1.2)$$

$$y_{1i} = \begin{cases} 1, & \text{if } y_{1i}^* > 0 \\ 0, & y_{1i}^* \leq 0. \end{cases} \quad (8.1.3)$$

$$y_{2i} = \begin{cases} y_{2i}^*, & \text{if } y_{1i}^* > 0 \\ 0, & y_{1i}^* \leq 0. \end{cases} \quad (8.1.4)$$

Assume

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & \sigma^2 \end{pmatrix}\right). \quad (8.1.5)$$

The Heckman two-step estimator in this case is as follows:  $\gamma$  can be consistently estimated with probit model. Further we may write:

$$\mathbb{E}[Y_2|Y_1 > 0, x, z] = x' \beta + \mathbb{E}[U_2|U_1 > -z' \gamma] = x' \beta + \rho \sigma \lambda(-z' \gamma) \quad (8.1.6)$$

$$\text{Var}[Y_2|Y_1 > 0, x, z] = \mathbb{E}[U_2^2|U_1 > -z' \gamma] = \sigma^2 + \rho^2 \sigma^2 [\lambda^2(z' \gamma) - \lambda(z' \gamma)^2] \quad (8.1.7)$$

where  $\lambda(x) = \phi(x)/\Phi(x)$ ;  $\Phi(\cdot)$  and  $\phi(\cdot)$  are the normal cumulative distribution function and density function respectively.  $\rho$  and  $\sigma$  can be estimated regressing  $y_{2i}$  on  $x_i$  and  $\lambda(-z' \gamma)$ . From the coefficient of the latter,  $\beta_\lambda$  and the residual variance  $s^2$ , one can isolate  $\rho$  and  $\sigma$ :

$$\hat{\sigma}^2 = s^2 + \beta_\lambda^2 [\lambda^2(z' \gamma) - \lambda(z' \gamma)^2] \quad (8.1.8)$$

$$\hat{\rho} = \frac{\beta_\lambda}{\hat{\sigma}}. \quad (8.1.9)$$

Note that  $\hat{\rho}$  need not to be in  $[-1, 1]$ .

Denote:

$$r = \sqrt{1 - \rho^2} \quad (8.1.10)$$

$$u_{2i} = y_{2i} - x_i' \beta \quad (8.1.11)$$

$$B_i = \frac{z_i' \gamma + \frac{\rho}{\sigma} u_{2i}}{r} \quad (8.1.12)$$

$$C(B) = -\frac{\Phi(B)\phi(B)B + \phi(B)^2}{\Phi(B)^2} \quad (8.1.13)$$

The contribution of observation  $i$  to the log-likelihood:

$$\ell = \sum_{i: y_{1i} \leq 0} \log \Phi(-z_i' \gamma) + \quad (8.1.14)$$



$$+ \sum_{i:y_{1i}>0} \left[ \log \Phi(B_i) - \frac{1}{2} \log 2\pi - \log \sigma - \frac{1}{2} \frac{u_{2i}^2}{\sigma^2} \right]. \quad (8.1.15)$$

The gradient of the log-likelihood is:

$$\frac{\partial \ell}{\partial \gamma} = \sum_{i:y_{1i} \leq 0} -\lambda(z'_i \gamma) z_i + \sum_{i:y_{1i} > 0} \lambda(B_i) \frac{z_i}{r} \quad (8.1.16)$$

$$\frac{\partial \ell}{\partial \beta} = \sum_{i:y_{1i} > 0} \left[ \frac{u_{2i}}{\sigma^2} - \lambda(B_i) \frac{\rho}{\sigma} \frac{1}{r} \right] x_i \quad (8.1.17)$$

$$\frac{\partial \ell}{\partial \sigma} = \sum_{i:y_{1i} > 0} \left[ \frac{u_{2i}^2}{\sigma^3} - \frac{1}{\sigma} - \lambda(B_i) \frac{\rho}{\sigma^2} \frac{u_{2i}}{r} \right] \quad (8.1.18)$$

$$\frac{\partial \ell}{\partial \rho} = \sum_{i:y_{1i} > 0} \lambda(B_i) \frac{\frac{1}{\sigma} u_{2i} + \rho z'_i \gamma}{r^3}. \quad (8.1.19)$$

Hessian components are

$$\frac{\partial^2 \ell}{\partial \gamma \gamma'} = - \sum_{i:y_{1i}=0} C(-z'_i \gamma) z_i z'_i + \sum_{i:y_{1i}=1} \frac{C(B)}{r} z_i z'_i \quad (8.1.20)$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \beta'} = - \sum_{i:y_{1i}=1} C(B) \frac{1}{\sigma} \frac{\rho}{r} z_i x'_i \quad (8.1.21)$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \sigma} = - \sum_{i:y_{1i}=1} C(B) \frac{\rho u_2}{\sigma^2 r^2} z_i \quad (8.1.22)$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \rho} = \sum_{i:y_{1i}=1} \left[ C(B) \frac{\frac{u_2}{\sigma} + \rho z'_i \gamma}{r^4} + \lambda(B) \frac{\rho}{r^3} \right] z_i \quad (8.1.23)$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta'} = \sum_{i:y_{1i}=1} \frac{1}{\sigma^2} \left[ \frac{\rho^2}{r^2} C(B) - 1 \right] x_i x'_i \quad (8.1.24)$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \sigma} = \sum_{i:y_{1i}=1} \left[ C(B) \frac{\rho^2}{\sigma^3} \frac{u_2}{r^2} + \frac{\rho}{\sigma^2} \frac{\lambda(B)}{r} - 2 \frac{u_2}{\sigma^3} \right] x_i \quad (8.1.25)$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \rho} = \sum_{i:y_{1i}=1} \left[ -C(B) \frac{\frac{u_2}{\sigma} + \rho z'_i \gamma}{r^4} \frac{\rho}{\sigma} - \frac{\lambda(B)}{\sigma} \frac{1}{r^3} \right] x_i \quad (8.1.26)$$

$$\frac{\partial^2 \ell}{\partial \sigma^2} = \sum_{i:y_{1i}=1} \left[ \frac{1}{\sigma^2} - 3 \frac{u_2^2}{\sigma^4} + 2 \lambda(B) \frac{u_2}{r} \frac{\rho}{\sigma^3} + \frac{\rho^2}{\sigma^4} \frac{u_2^2}{r^2} C(B) \right] \quad (8.1.27)$$

$$\frac{\partial^2 \ell}{\partial \sigma \partial \rho} = -\frac{1}{r^3} \sum_{i:y_{1i}=1} \frac{u_2}{\sigma^2} \left[ C(B) \frac{\rho \left( \frac{u_2}{\sigma} + \rho z'_i \gamma \right)}{r} + \lambda(B) \right] \quad (8.1.28)$$

$$\frac{\partial^2 \ell}{\partial \varrho^2} = \sum_{i: y_{1i}=1} \left[ C(B) \left( \frac{\frac{u_2}{\sigma} + \varrho z'_i \gamma}{r^3} \right)^2 + \lambda(B) \frac{z'_i \gamma (1 + 2\varrho^2) + 3\varrho \frac{u_2}{\sigma}}{r^5} \right] \quad (8.1.29)$$

## 8.2 Tobit-2 Model with Binary Outcome

The underlying latent model:

$$y_i^{S*} = \beta^{S'} x_i^S + \epsilon_i^S \quad (8.2.1)$$

$$y_i^{O*} = \beta^{O'} x_i^O + \epsilon_i^O \quad (8.2.2)$$

$$y_i^S = \begin{cases} 1, & \text{if } y_i^{S*} > 0 \\ 0, & y_i^{S*} \leq 0. \end{cases} \quad (8.2.3)$$

$$y_i^O = \begin{cases} \text{undetermined,} & \text{if } y_i^S = 0 \quad (\text{case 1}) \\ 0, & \text{if } y_i^{O*} \leq 0 \quad \text{and } y_i^S = 1 \quad (\text{case 2}) \\ 1, & \text{if } y_i^{O*} > 0 \quad \text{and } y_i^S = 1 \quad (\text{case 3}) \end{cases} \quad (8.2.4)$$

Assume

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix} \right). \quad (8.2.5)$$

The log-likelihood function contains 3 components, corresponding to the cases in (8.2.4):

$$\ell = \sum_{i \in \text{case 1}} \log \Phi(-\beta^{S'} x_i^S) \quad (8.2.6)$$

$$+ \sum_{i \in \text{case 2}} \log \left[ 1 - \Phi(-\beta^{S'} x_i^S) - \bar{\Phi}_2 \left( \begin{pmatrix} -\beta^{S'} x_i^S \\ -\beta^{O'} x_i^O \end{pmatrix}, \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix} \right) \right] \quad (8.2.7)$$

$$+ \sum_{i \in \text{case 3}} \log \bar{\Phi}_2 \left( \begin{pmatrix} -\beta^{S'} x_i^S \\ -\beta^{O'} x_i^O \end{pmatrix}, \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix} \right), \quad (8.2.8)$$

where  $\bar{\Phi}_2(\cdot, \cdot)$  is the upper tail probability of 2-dimensional normal distribution.

Denote by  $\mathcal{L}_i$  the corresponding individual likelihood value. The score vector:

$$\begin{aligned} \frac{\partial}{\partial \beta^S} \ell &= \sum_{i \in \text{case 1}} \frac{1}{\mathcal{L}_i} \phi(-\beta^{S'} x_i^S) x_i^S \\ &+ \sum_{i \in \text{case 2}} \frac{1}{\mathcal{L}_i} \phi(\beta^{S'} x_i^S) \bar{\Phi} \left( \frac{\beta^{O'} x_i^O - \varrho \beta^{S'} x_i^S}{\sqrt{1 - \varrho^2}} \right) x_i^S \\ &+ \sum_{i \in \text{case 3}} \frac{1}{\mathcal{L}_i} \phi(\beta^{S'} x_i^S) \Phi \left( \frac{\beta^{O'} x_i^O - \varrho \beta^{S'} x_i^S}{\sqrt{1 - \varrho^2}} \right) x_i^S \\ \frac{\partial}{\partial \beta^O} \ell &= \sum_{i \in \text{case 2}} \frac{1}{\mathcal{L}_i} \phi(\beta^{O'} x_i^O) \bar{\Phi} \left( \frac{\beta^{S'} x_i^S - \varrho \beta^{O'} x_i^O}{\sqrt{1 - \varrho^2}} \right) x_i^O \\ &+ \sum_{i \in \text{case 3}} \frac{1}{\mathcal{L}_i} \phi(\beta^{O'} x_i^O) \Phi \left( \frac{\beta^{S'} x_i^S - \varrho \beta^{O'} x_i^O}{\sqrt{1 - \varrho^2}} \right) x_i^O \end{aligned}$$

$$\frac{\partial}{\partial \varrho} \ell = - \sum_{i \in \text{case 2}} \phi_2 \left( \begin{pmatrix} -\boldsymbol{\beta}^{S'} x_i^S \\ -\boldsymbol{\beta}^{O'} x_i^O \end{pmatrix}, \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix} \right) + \sum_{i \in \text{case 3}} \phi_2 \left( \begin{pmatrix} -\boldsymbol{\beta}^{S'} x_i^S \\ -\boldsymbol{\beta}^{O'} x_i^O \end{pmatrix}, \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix} \right),$$

where  $\phi_2(\cdot, \cdot)$  is 2-dimensional normal density.

### 8.3 Tobit-5 Model

Definitsioon (lühiduse mõttes on indeks  $i$  ära jäetud):

$$y_1^* = \mathbf{Z}'\boldsymbol{\gamma} + u_1 \quad (8.3.1)$$

$$y_2^* = \mathbf{X}'\boldsymbol{\beta}_2 + u_2 \quad (8.3.2)$$

$$y_3^* = \mathbf{X}'\boldsymbol{\beta}_3 + u_3 \quad (8.3.3)$$

$$y_2 = \begin{cases} y_2^* & \text{kui } y_1^* \leq 0 \\ 0 & \text{kui } y_1^* > 0 \end{cases} \quad (8.3.4)$$

$$y_3 = \begin{cases} y_3^* & \text{kui } y_1^* > 0 \\ 0 & \text{kui } y_1^* \leq 0 \end{cases} \quad (8.3.5)$$

Eeldatakse et jääkliikmete jaotus on niisugune:

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \varrho_2\sigma_2 & \varrho_3\sigma_3 \\ \varrho_2\sigma_2 & \sigma_2^2 & \sigma_{23} \\ \varrho_3\sigma_3 & \sigma_{23} & \sigma_3^2 \end{pmatrix} \right). \quad (8.3.6)$$

#### 8.3.1 Heckmani kahesammuline hinnang

$\hat{\gamma}$  leitakse probiti abil. Edasi võib kirjutada

$$\begin{aligned} y_2 &= \mathbf{X}'\boldsymbol{\beta}_2 - \varrho_2\sigma_2\lambda(-\mathbf{Z}'\boldsymbol{\gamma}) + e_2 \\ y_3 &= \mathbf{X}'\boldsymbol{\beta}_3 + \varrho_3\sigma_3\lambda(\mathbf{Z}'\boldsymbol{\gamma}) + e_3 \end{aligned} \quad (8.3.7)$$

Kusjuures

$$\begin{aligned} \sigma_{e2} &= \sigma_2^2 \left\{ 1 - \varrho_2^2 \left[ \lambda^2(-\mathbf{Z}'\boldsymbol{\gamma}) - \mathbf{Z}'\boldsymbol{\gamma}\lambda(-\mathbf{Z}'\boldsymbol{\gamma}) \right] \right\} \\ \sigma_{e3} &= \sigma_3^2 \left\{ 1 - \varrho_3^2 \left[ \lambda^2(\mathbf{Z}'\boldsymbol{\gamma}) + \mathbf{Z}'\boldsymbol{\gamma}\lambda(\mathbf{Z}'\boldsymbol{\gamma}) \right] \right\} \end{aligned} \quad (8.3.8)$$

Kui lähendada seost (8.3.7) OLS-ga, siis saab  $\lambda$  koefitsendi ja dispersiooni hinnangu abil leida  $\hat{\varrho}$  ja  $\hat{\sigma}$ . Märkus:  $\hat{\varrho}$  ei pruugi olla -1 ja 1 vahel.

#### 8.3.2 Maksimum-laiklikhuud hinnang

Mudeli log-laiklikhuud on:

$$\begin{aligned} l &= -\frac{N}{2} \log 2\pi + \\ &+ \sum_{i \in \text{case } 2} \left\{ -\log \sigma_2 - \frac{1}{2} \left( \frac{u_2}{\sigma_2} \right)^2 + \log \Phi \left[ \frac{\mathbf{Z}'\boldsymbol{\gamma} + \frac{\varrho_2}{\sigma_2} (y_2 - \mathbf{X}'_i\boldsymbol{\beta}_2)}{\sqrt{1 - \varrho_2^2}} \right] \right\} \\ &+ \sum_{i \in \text{case } 3} \left\{ -\log \sigma_3 - \frac{1}{2} \left( \frac{y_3 - \mathbf{X}'_i\boldsymbol{\beta}_3}{\sigma_3} \right)^2 + \log \Phi \left[ \frac{\mathbf{Z}'\boldsymbol{\gamma} + \frac{\varrho_3}{\sigma_3} (y_3 - \mathbf{X}'_i\boldsymbol{\beta}_3)}{\sqrt{1 - \varrho_3^2}} \right] \right\}. \end{aligned} \quad (8.3.9)$$

Valikud 2 ja 3 erinevad ainult avaldise märgi poolest funktsiooni  $\Phi$  sees. Tuletised on:

$$\frac{\partial l}{\partial \gamma} = -\sum_2 \frac{\phi(B_2)}{\Phi(B_2)} \frac{\mathbf{Z}}{\sqrt{1-\varrho_2^2}} + \sum_3 \frac{\phi(B_3)}{\Phi(B_3)} \frac{\mathbf{Z}}{\sqrt{1-\varrho_3^2}} \quad (8.3.10)$$

$$\frac{\partial l}{\partial \beta_2} = \sum_2 \left[ \frac{\phi(B_2)}{\Phi(B_2)} \left( \frac{\varrho_2}{\sigma_2} \frac{\mathbf{X}}{\sqrt{1-\varrho_2^2}} \right) + \frac{u_2}{\sigma_2^2} \mathbf{X} \right] \quad (8.3.11)$$

$$\frac{\partial l}{\partial \sigma_2} = \sum_2 \left[ -\frac{1}{\sigma_2} + \frac{(y_2 - \mathbf{X}'\beta_2)^2}{\sigma_2^3} + \frac{\phi(B_2)}{\Phi(B_2)} \frac{\varrho_2}{\sigma_2^2} \frac{y_2 - \mathbf{X}'\beta_2}{\sqrt{1-\varrho_2^2}} \right] \quad (8.3.12)$$

$$\frac{\partial l}{\partial \varrho_2} = -\sum_2 \frac{\phi(B_2)}{\Phi(B_2)} \frac{\frac{1}{\sigma_2} (y_2 - \mathbf{X}'\beta_2) + \varrho_2 \mathbf{Z}'\gamma}{(1-\varrho_2^2)^{\frac{3}{2}}} \quad (8.3.13)$$

$$\frac{\partial l}{\partial \beta_3} = \sum_3 \left[ -\frac{\phi(B_3)}{\Phi(B_3)} \left( \frac{\varrho_3}{\sigma_3} \frac{\mathbf{X}}{\sqrt{1-\varrho_3^2}} \right) + \frac{u_3}{\sigma_3^2} \mathbf{X} \right] \quad (8.3.14)$$

$$\frac{\partial l}{\partial \sigma_3} = \sum_3 \left[ -\frac{1}{\sigma_3} + \frac{(y_3 - \mathbf{X}'\beta_3)^2}{\sigma_3^3} - \frac{\phi(B_3)}{\Phi(B_3)} \frac{\varrho_3}{\sigma_3^2} \frac{y_3 - \mathbf{X}'\beta_3}{\sqrt{1-\varrho_3^2}} \right] \quad (8.3.15)$$

$$\frac{\partial l}{\partial \varrho_3} = \sum_3 \frac{\phi(B_3)}{\Phi(B_3)} \frac{\frac{1}{\sigma_3} (y_3 - \mathbf{X}'\beta_3) + \varrho_3 \mathbf{Z}'\gamma}{(1-\varrho_3^2)^{\frac{3}{2}}} \quad (8.3.16)$$

Teised tuletised:

$$\frac{\partial^2 l}{\partial \gamma^2} = \sum_2 \frac{C(B_2)}{1-\varrho_2^2} \mathbf{z}_i \mathbf{z}_i' + \sum_3 \frac{C(B_3)}{1-\varrho_3^2} \mathbf{z}_i \mathbf{z}_i' \quad (8.3.17)$$

$$\frac{\partial^2 l}{\partial \gamma \partial \beta_2'} = -\sum_2 C(B_2) \frac{1}{\sigma_2} \frac{\varrho_2}{1-\varrho_2^2} \mathbf{Z} \mathbf{X}' \quad (8.3.18)$$

$$\frac{\partial^2 l}{\partial \gamma \partial \sigma_2} = -\sum_2 \frac{\varrho_2 u_2}{\sigma_2^2 (1-\varrho_2^2)} C(B_2) \mathbf{Z} \quad (8.3.19)$$

$$\frac{\partial^2 l}{\partial \gamma \partial \varrho_2} = \sum_2 \left[ C(B_2) \frac{\frac{u_2}{\sigma_2} \varrho_2 \mathbf{Z}'\gamma}{(1-\varrho_2^2)^2} - \lambda(B_2) \frac{\varrho_2}{(1-\varrho_2^2)^{\frac{3}{2}}} \right] \mathbf{Z} \quad (8.3.20)$$

$$\frac{\partial^2 l}{\partial \gamma \partial \beta_3'} = -\sum_3 C(B_3) \frac{1}{\sigma_3} \frac{\varrho_3}{1-\varrho_3^2} \mathbf{Z} \mathbf{X}' \quad (8.3.21)$$

$$\frac{\partial^2 l}{\partial \gamma \partial \sigma_3} = -\sum_3 \frac{\varrho_3 u_3}{\sigma_3^2 (1-\varrho_3^2)} C(B_3) \mathbf{Z} \quad (8.3.22)$$

$$\frac{\partial^2 l}{\partial \gamma \partial \varrho_3} = \sum_3 \left[ C(B_3) \frac{\frac{u_3}{\sigma_3} \varrho_3 \mathbf{Z}' \gamma}{(1 - \varrho_3^2)^2} + \lambda(B_3) \frac{\varrho_3}{(1 - \varrho_3^2)^{\frac{3}{2}}} \right] \mathbf{Z} \quad (8.3.23)$$

$$\frac{\partial^2 l}{\partial \beta_2 \partial \beta'_2} = \sum_2 \frac{1}{\sigma_2^2} \left[ \frac{\varrho_2^2}{1 - \varrho_2^2} C(B_2) - 1 \right] \mathbf{X} \mathbf{X}' \quad (8.3.24)$$

$$\frac{\partial^2 l}{\partial \beta_2 \partial \sigma_2} = \sum_2 \left[ C(B_2) \frac{u_2}{\sigma_2^3} \frac{\varrho_2^2}{1 - \varrho_2^2} - \frac{\lambda(B_2)}{\sigma_2^2} \frac{\varrho_2}{\sqrt{1 - \varrho_2^2}} - 2 \frac{u_2}{\sigma_2^3} \right] \mathbf{X} \quad (8.3.25)$$

$$\frac{\partial^2 l}{\partial \beta_2 \partial \varrho_2} = \sum_2 \left[ -C(B_2) \frac{\frac{u_2}{\sigma_2} + \varrho_2 \mathbf{Z}' \gamma}{(1 - \varrho_2^2)^2} \frac{\varrho_2}{\sigma_2} + \frac{\lambda(B_2)}{\sigma_2} \frac{1}{(1 - \varrho_2^2)^{\frac{3}{2}}} \right] \mathbf{X} \quad (8.3.26)$$

$$\frac{\partial^2 l}{\partial \beta_2 \partial \beta_3} = 0 \quad (8.3.27)$$

$$\frac{\partial^2 l}{\partial \beta_2 \partial \sigma_3} = 0 \quad (8.3.28)$$

$$\frac{\partial^2 l}{\partial \beta_2 \partial \varrho_3} = 0 \quad (8.3.29)$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \sigma_2^2} &= \sum_2 \left[ \frac{1}{\sigma_2^2} - 3 \frac{u_2^2}{\sigma_2^4} + \frac{u_2}{\sigma_2^4} \frac{\varrho_2^2}{1 - \varrho_2^2} C(B_2) \right] - \\ &- 2 \sum_2 \lambda(B_2) \frac{u_2}{\sigma_2^3} \frac{\varrho_2}{\sqrt{1 - \varrho_2^2}} \end{aligned} \quad (8.3.30)$$

$$\frac{\partial^2 l}{\partial \sigma_2 \partial \varrho_2} = \frac{1}{(1 - \varrho_2^2)^{\frac{3}{2}}} \sum_2 \frac{u_2}{\sigma_2^2} \left[ -C(B_2) \frac{\varrho_2 \left( \frac{u_2}{\sigma_2} + \varrho_2 \mathbf{Z}' \gamma \right)}{\sqrt{1 - \varrho_2^2}} + \lambda(B_2) \right] \quad (8.3.31)$$

$$\frac{\partial^2 l}{\partial \sigma_2 \partial \beta_3} = 0 \quad (8.3.32)$$

$$\frac{\partial^2 l}{\partial \sigma_2 \partial \sigma_3} = 0 \quad (8.3.33)$$

$$\frac{\partial^2 l}{\partial \sigma_2 \partial \varrho_3} = 0 \quad (8.3.34)$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \varrho_2^2} &= \sum_2 C(B_2) \left[ \frac{\frac{u_2}{\sigma_2} + \varrho_2 \mathbf{Z}' \gamma}{(1 - \varrho_2^2)^{\frac{3}{2}}} \right]^2 - \\ &- \sum_2 \frac{\phi(B_2)}{\Phi(B_2)} \frac{\mathbf{Z}' \gamma (1 + 2\varrho_2^2) + 3\varrho_2 \frac{u_2}{\sigma_2}}{(1 - \varrho_2^2)^{\frac{5}{2}}} \end{aligned} \quad (8.3.35)$$

$$\frac{\partial^2 l}{\partial \varrho_2 \partial \beta_3} = 0 \quad (8.3.36)$$

$$\frac{\partial^2 l}{\partial \varrho_2 \partial \sigma_3} = 0 \quad (8.3.37)$$

$$\frac{\partial^2 l}{\partial \varrho_2 \partial \varrho_3} = 0 \quad (8.3.38)$$

$$\frac{\partial^2 l}{\partial \beta_3 \partial \beta'_3} = \sum_3 \frac{1}{\sigma_3^2} \left[ \frac{\varrho_3^2}{1 - \varrho_3^2} C(B_3) - 1 \right] \mathbf{X} \mathbf{X}' \quad (8.3.39)$$

$$\frac{\partial^2 l}{\partial \beta_3 \partial \sigma_3} = \sum_3 \left[ C(B_3) \frac{\varrho_3^2}{\sigma_3^3} \frac{u_3}{1 - \varrho_3^2} + \frac{\varrho_3}{\sigma_3^2} \frac{\lambda(B_3)}{\sqrt{1 - \varrho_3^2}} - 2 \frac{u_3}{\sigma_3^3} \right] \mathbf{X} \quad (8.3.40)$$

$$\frac{\partial^2 l}{\partial \beta_3 \partial \varrho_3} = \sum_3 \left[ -C(B_3) \frac{\frac{u_3}{\sigma_3} + \varrho_3 \mathbf{Z}' \boldsymbol{\gamma}}{(1 - \varrho_3^2)^2} \frac{\varrho_3}{\sigma_3} - \frac{\lambda(B_3)}{\sigma_3} \frac{1}{(1 - \varrho_3^2)^{\frac{3}{2}}} \right] \mathbf{X} \quad (8.3.41)$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \sigma_3^2} &= \sum_3 \left[ \frac{1}{\sigma_3^2} - 3 \frac{u_3^2}{\sigma_3^4} + 2 \lambda(B_3) \frac{y_3 - \mathbf{X}' \boldsymbol{\beta}_3}{\sqrt{1 - \varrho_3^2}} \frac{\varrho_3}{\sigma_3^3} \right] + \\ &+ \sum_3 \frac{\varrho_3^2}{\sigma_3^4} \frac{u_3^2}{1 - \varrho_3^2} C(B_3) \end{aligned} \quad (8.3.42)$$

$$\frac{\partial^2 l}{\partial \sigma_3 \partial \varrho_3} = -\frac{1}{(1 - \varrho_3^2)^{\frac{3}{2}}} \sum_3 \frac{u_3}{\sigma_3^2} \left[ C(B_3) \frac{\varrho_3 \left( \frac{u_3}{\sigma_3} + \varrho_3 \mathbf{Z}' \boldsymbol{\gamma} \right)}{\sqrt{1 - \varrho_3^2}} + \lambda(B_3) \right] \quad (8.3.43)$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \varrho_3^2} &= \sum_3 C(B_3) \left[ \frac{\frac{1}{\sigma_3} u_3 + \varrho_3 \mathbf{Z}' \boldsymbol{\gamma}}{(1 - \varrho_3^2)^{\frac{3}{2}}} \right]^2 + \\ &+ \sum_3 \lambda(B_3) \frac{\mathbf{Z}' \boldsymbol{\gamma} (1 + 2\varrho_3^2) + 3\varrho_3 \frac{1}{\sigma_3} u_3}{(1 - \varrho_3^2)^{\frac{7}{2}}} \end{aligned} \quad (8.3.44)$$

Siin on tähistatud

$$B_2 = -\frac{\mathbf{Z}' \boldsymbol{\gamma} + \frac{\varrho_2}{\sigma_2} (y_2 - \mathbf{X}' \boldsymbol{\beta}_2)}{\sqrt{1 - \varrho_2^2}} \quad (8.3.45)$$

$$B_3 = \frac{\mathbf{Z}' \boldsymbol{\gamma} + \frac{\varrho_3}{\sigma_3} (y_3 - \mathbf{X}' \boldsymbol{\beta}_3)}{\sqrt{1 - \varrho_3^2}} \quad (8.3.46)$$

$$\lambda(B) = \frac{\phi(B)}{\Phi(B)} \quad (8.3.47)$$

$$u_2 = y_2 - \mathbf{X}' \boldsymbol{\beta}_2 \quad (8.3.48)$$

$$u_3 = y_3 - \mathbf{X}' \boldsymbol{\beta}_3 \quad (8.3.49)$$

$$C(B) = -\frac{\Phi(B)\phi(B)B + \phi(B)^2}{\Phi(B)^2} \quad (8.3.50)$$



## 8.4 Kestusmodelid

Tähistused:

$\tau$  kestus, algseisundis viibitud aeg

$t$  kalendriaeg

### 8.4.1 Kaplan-Meieri hinnang

**KM hinnang diskreetses ajas** Olgu perioodil  $j$   $r_j$  inimest “riski hulgas”, s.t.  $r_j$  inimest võiksid põhimõtteliselt seisundist lahkuda. Lahkugu tegelikult  $n_j$  inimest,  $r_j - n_j$  jäävad edasi algseisundisse. KM hinnang hasardile on seega

$$\hat{h} = \frac{n_j}{r_j} \quad (8.4.1)$$

$$\widehat{\text{Var}} \hat{h}_j = \frac{\hat{h}_j(1 - \hat{h}_j)}{r_j}. \quad (8.4.2)$$

Kui periood  $j$  on  $k$  kuu pikkune, siis (keskmise) ühe kuu spetsiifilise hasardi saab

$$\hat{\vartheta}_j = 1 - (1 - \hat{h}_j)^{1/k} \quad (8.4.3)$$

$$\widehat{\text{Var}} \hat{\vartheta}_j = \frac{\text{Var} \hat{h}_j}{\left[k(1 - \hat{h}_j)^{1-1/k}\right]^2} \quad (8.4.4)$$

**KM hinnang pidevas ajas** Lahkugu aja  $t$  jooksul  $r$  algseisundis olnud inimestest  $n$ . Keskmise hasart ajaühikus on

$$\hat{\vartheta} = -\frac{1}{t} \log(1 - n/r) \quad (8.4.5)$$

$$\widehat{\text{Var}} \hat{\vartheta} = \frac{1}{t^2} \frac{n/r}{r - n}. \quad (8.4.6)$$

### 8.4.2 Multiplikatiivne mittevaadeldav heterogeensus

Eeldame et hasart avaldub

$$\vartheta(\tau|x, v) = \lambda(\tau|x)v \quad (8.4.7)$$

kus  $v$  on mingi kindla jaotusega mittevaadeldav juhuslik suurus. Nüüd  $v$  keskvärtus algseisundisse jääjatel sõltub ajast:

$$\mathbb{E}(v|T \geq \tau) = -\frac{\mathcal{L}'[z(\tau|x)]}{\mathcal{L}[z(\tau|x)]} \quad (8.4.8)$$

kus on integreeritud hasart  $v$ -d arvestamata:

$$z(\tau|x) = \int_0^\tau \lambda(s|x) ds \quad (8.4.9)$$

Kui  $v$  on algseisundisse sissevoolus ühikdispersiooniga gammajaotus parameetriga  $\alpha$ , siis

$$\mathbb{E}(v|T \geq \tau) = \frac{\alpha}{z(\tau|x) + \alpha^{1/2}}. \quad (8.4.10)$$

### 8.4.3 Tükati konstantne põhihasart ja diskreetne mittevaadeldav heterogeensus ning pidev aeg

**Sõltumatud vaatlused** Eeldatakse, et hasart on konstantne iga  $M$  ajavahemiku sees, erinevatel ajavahemikel võib ta aga olla erinev. Olgu hasart kirjeldatud vektoriga  $\lambda$ , kusjuures ajavahemiku  $j$  põhihasart olgu  $e^{\lambda_j}$ . Mittevaadeldav heterogeensus on diskreetse jaotusega:

$$v = \begin{cases} v_h \equiv 1, & \text{tõenäosusega } p_h, \\ v_l, & \text{tõenäosusega } p_l = 1 - p_h. \end{cases} \quad (8.4.11)$$

Sobiv on parametrizeerida

$$\begin{aligned} v_1 &= e^{\tilde{v}_1} & p_1 &= \Lambda(\tilde{p}_1) \\ \dots & & \text{ja} & \dots \\ v_K &= 1 & p_K &= 1 - \sum^{K-1} p_k, \end{aligned} \quad (8.4.12)$$

Kus  $\Lambda(\cdot)$  on logistile jaotusfunktsioon ja  $v_k \in \mathfrak{R}$  ning  $p_k \in \mathfrak{R}$ .

Olgu  $m_i$  ajaperiood, mille jooksul inimene lahkub uuritavast seisundist ja tsenseerimist kirjeldagu  $\delta_i = 0$  kui vaatlus on tsenseeritud ja 1 kui tsenseerimata ning  $\mu_i = e^{\gamma \lambda_i}$  olgu hasardi inimesest sõltuv osa.  $T_{ij}$  olgu teadaolev (võimalik et tsenseeritud) aeg, mis inimene  $i$  veetis uuritavas seisundis ajaperioodi  $j$  jooksul. Vektor  $T_i$  on  $M$ -vektor, mille komponendid on  $T_{ij}$  ning vektor  $d_i$  on vektor, mille  $j$ -s komponent on 1, kui inimene lahkus uuritavast seisundist ajaperioodil  $j$ . Muud komponendid on nullid.

Sel juhul inimese  $i$  laiklihuud avaldub:

$$\mathcal{L}_i = p_l \mathcal{L}_{li} + p_h \mathcal{L}_{hi} = p_l \left( v_l \mu_i e^{\lambda_{m_i}} \right)^{\delta_i} e^{-z_{li}} + p_h \left( \mu_i e^{\lambda_{m_i}} \right)^{\delta_i} e^{-z_{hi}}, \quad (8.4.13)$$

kus

$$z_{li} = v_l \mu_i \sum_{j=1}^{M-1} e^{\lambda_j} T_{ij} \quad \text{ja} \quad z_{hi} = \mu_i \sum_{j=1}^{M-1} e^{\lambda_j} T_{ij} \quad (8.4.14)$$

on integreeritud hasart. Log-laiklihuudi gradient avaldub:

$$\frac{\partial \ell_i}{\partial v_l} = \frac{p_l \mathcal{L}_{li}}{\mathcal{L}_i} \left[ \frac{\delta_i}{v_l} - z_{hi} \right] \quad (8.4.15)$$

$$\frac{\partial \ell_i}{\partial p_l} = \frac{1}{\mathcal{L}_i} [\mathcal{L}_{li} - \mathcal{L}_{hi}] \quad (8.4.16)$$

$$\frac{\partial \ell_i}{\partial \lambda} = p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i} (\delta_i d_i - v_l \mu_i T_i) + p_h \frac{\mathcal{L}_{hi}}{\mathcal{L}_i} (\delta_i d_i - \mu_i T_i) \quad (8.4.17)$$

$$\frac{\partial \ell_i}{\partial \gamma} = \frac{x_i}{\mathcal{L}_i} [p_l \mathcal{L}_l (\delta_i - z_{li}) + p_h \mathcal{L}_h (\delta_i - z_{hi})] \quad (8.4.18)$$

$$(8.4.19)$$

ja hessi maatriks:

$$\frac{\partial^2 \ell_i}{\partial v_l^2} = \frac{\mathcal{L}_{li}}{\mathcal{L}_i} \left[ \left( \frac{\delta_i}{v_l} - z_{hi} \right)^2 \left( 1 - p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i} \right) - \frac{\delta_i}{v_l^2} \right] \quad (8.4.20)$$

$$\frac{\partial^2 \ell_i}{\partial p_l^2} = -\left(\frac{\mathcal{L}_{li}}{\mathcal{L}_i} - \frac{\mathcal{L}_{hi}}{\mathcal{L}_i}\right)^2 \quad (8.4.21)$$

$$\frac{\partial^2 \ell_i}{\partial v_l \partial p_l} = \frac{\mathcal{L}_{li}}{\mathcal{L}_i} \left(\frac{\delta_i}{v_l} - z\right) \left(1 - p_l \frac{\mathcal{L}_{li} - \mathcal{L}_{hi}}{\mathcal{L}_i}\right) \quad (8.4.22)$$

$$\begin{aligned} \frac{\partial^2 \ell_i}{\partial \lambda \partial \lambda'} &= \frac{p_l}{\mathcal{L}_i} \left[ \frac{\partial \mathcal{L}_{li}}{\partial \lambda} (\delta_i \mathbf{d}_i - v_l \mu_i \mathbf{T}_i) - \mathcal{L}_{li} \text{diag}(v_l \mu_i \mathbf{e}^\lambda * \mathbf{T}_i) \right] + \\ &+ \frac{p_h}{\mathcal{L}_i} \left[ \frac{\partial \mathcal{L}_{hi}}{\partial \lambda} (\delta_i \mathbf{d}_i - \mu_i \mathbf{T}_i) - \mathcal{L}_{hi} \text{diag}(\mu_i \mathbf{e}^\lambda * \mathbf{T}_i) \right] - \\ &- \frac{1}{\mathcal{L}_i^2} \left( \frac{\partial \mathcal{L}_i}{\partial \lambda} \right)^2 \end{aligned} \quad (8.4.23)$$

$$\begin{aligned} \frac{\partial^2 \ell_i}{\partial \gamma \partial \gamma'} &= p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i} [(\delta_i - z_{li})^2 - z_{li}] \mathbf{x}_i \mathbf{x}_i' + p_h \frac{\mathcal{L}_{hi}}{\mathcal{L}_i} [(\delta_i - z_{hi})^2 - z_{hi}] \mathbf{x}_i \mathbf{x}_i' - \\ &- \frac{\partial \mathcal{L}_i}{\partial \gamma} \frac{\partial \mathcal{L}_i}{\partial \gamma'} \end{aligned} \quad (8.4.24)$$

$$\begin{aligned} \frac{\partial^2 \ell_i}{\partial \lambda \gamma'} &= [(\delta_i - z_{li})(\delta_i \mathbf{d}_i - z_{li}) - z_{li}] \frac{\mathcal{L}_{li}}{\mathcal{L}_i} \mathbf{x}_i' p_l + \\ &+ [(\delta_i - z_{hi})(\delta_i \mathbf{d}_i - z_{hi}) - z_{hi}] \frac{\mathcal{L}_{hi}}{\mathcal{L}_i} \mathbf{x}_i' p_h - \frac{1}{\mathcal{L}_i} \frac{\partial \mathcal{L}_i}{\partial \lambda} \frac{\partial \mathcal{L}_i}{\partial \gamma'} \end{aligned} \quad (8.4.25)$$

$$\begin{aligned} \frac{\partial^2 \ell_i}{\partial \lambda \partial v_l} &= p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i} \left(\frac{\delta_i}{v_l} - z_{hi}\right) \left(\delta_i \mathbf{d}_i - v_l \mu_i \mathbf{e}^\lambda * \mathbf{T}_i - \frac{1}{\mathcal{L}_i} \frac{\partial \mathcal{L}_i}{\partial \lambda}\right) - \\ &- p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i} \mu_i \mathbf{e}^\lambda * \mathbf{T}_i \end{aligned} \quad (8.4.26)$$

$$\begin{aligned} \frac{\partial^2 \ell_i}{\partial \lambda \partial p_l} &= \frac{\mathcal{L}_{li}}{\mathcal{L}_i} (\delta_i \mathbf{d}_i - v_l \mu_i \mathbf{e}^\lambda * \mathbf{T}_i) - \frac{\mathcal{L}_{hi}}{\mathcal{L}_i} (\delta_i \mathbf{d}_i - \mu_i \mathbf{e}^\lambda * \mathbf{T}_i) - \\ &- \frac{\mathcal{L}_{li} - \mathcal{L}_{hi}}{\mathcal{L}_i} \frac{1}{\mathcal{L}_i} \frac{\partial \mathcal{L}_i}{\partial \lambda} \end{aligned} \quad (8.4.27)$$

$$\begin{aligned} \frac{\partial^2 \ell_i}{\partial \gamma \partial v_l} &= p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i} \left(\frac{\delta_i}{v_l} - z_{hi}\right) (\delta_i - z_{li}) \mathbf{x}_i - \\ &- p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i} \left(\frac{\delta_i}{v_l} - z_{hi}\right) \frac{1}{\mathcal{L}_i} \frac{\partial \mathcal{L}_i}{\partial \gamma} \mathbf{x}_i - p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i} z_i \mathbf{x}_i \end{aligned} \quad (8.4.28)$$

$$\frac{\partial^2 \ell_i}{\partial \gamma \partial p_l} = \left[ \frac{\mathcal{L}_{li}}{\mathcal{L}_i} (\delta_i - z_{li}) - \frac{\mathcal{L}_{hi}}{\mathcal{L}_i} (\delta_i - z_{hi}) \right] \mathbf{x}_i - \frac{\mathcal{L}_{li} - \mathcal{L}_{hi}}{\mathcal{L}_i} \frac{1}{\mathcal{L}_i} \frac{\partial \mathcal{L}_i}{\partial \gamma} \quad (8.4.29)$$

Eelnevas tähendab  $\{\mathbf{e}^\lambda\}_i = \mathbf{e}^{\lambda_i}$ ,  $*$  vektorite elementide kaupa korrutamist ( $\{\mathbf{a} * \mathbf{b}\}_i = a_i b_i$ ) ning  $\text{diag } \mathbf{a}$  on maatriks, mille peadiagonaalil on vektor  $\mathbf{a}$  ja mujal nullid.

**Indiviidi-spetsiifiline heterogeensus** Eeldame nii nagu eespool et hasart on konstantne iga  $M$  ajavahemiku sees, erinevatel ajavahemikel võib ta aga olla erinev. Avaldugu hasart

$$\theta(t|x, v) = v e^{\lambda(t)} e^{\gamma x'_i}. \quad (8.4.30)$$

Mittevaadeldav heterogeensus olgu diskreetse jaotusega  $v \in \{v_1, v_2, \dots, v_K\}$  ja tõenäosusega vastavalt  $p_1, p_2, \dots, p_K$ . Olgu inimese mittevadeldav tunnus  $v$  ajas muutumatu, vaadeldav tunnus aga võib muutuda. Inimese  $i$  spelli  $j$  osa laiklihuudi funktsioonis on siis:

$$\mathcal{L}_{ij}(\cdot|v) = v \theta(t_{ij}|x_{ij})^{\delta_{ij}} S(t_{ij}|x_{ij}, v). \quad (8.4.31)$$

Siin  $t_{ij}$  on spelli vaadeldud kestus,  $\delta_{ij}$  on mitte-tsenseerituse indikaator ja  $x_{ij}$  on inimese  $i$  vaadeldavad isikutunnused spelli  $j$  ajal.  $\mathcal{L}_{ij}(\cdot|v)$  on analoogne indiviidi-spetsiifilise laiklihuudiga sõltumatute vaatluste juhul.

Olgu inimese  $i$  kohta  $N_i$  vaadeldud spelli. Inimese  $i$  osa laiklihuudis avaldub siis

$$\mathcal{L}_i(\cdot|v) = \prod_{j=1}^{N_i} \mathcal{L}_{ij}(\cdot|v). \quad (8.4.32)$$

Vaadeldav laiklihuud avaldub:

$$\mathcal{L}_i(\cdot) = \sum_{k=1}^K p_k \mathcal{L}_i(\cdot|v_k). \quad (8.4.33)$$

Log-laiklihuudi gradient avaldub:

$$\frac{\partial \ell_i}{\partial v_k} = \frac{p_k}{\mathcal{L}_i(\cdot)} \sum_j \frac{\mathcal{L}_i(\cdot|v_k)}{\mathcal{L}_{ij}(\cdot|v_k)} \frac{\partial \mathcal{L}_{ij}(\cdot|v_k)}{\partial v_k} \quad (8.4.34)$$

$$\frac{\partial \ell_i}{\partial p_l} = \frac{\mathcal{L}_i(\cdot|v_l)}{\mathcal{L}_i(\cdot)} \quad (8.4.35)$$

$$\frac{\partial \ell_i}{\partial \lambda} = \frac{1}{\mathcal{L}_i(\cdot)} \sum_k p_k \sum_j \frac{\mathcal{L}_i(\cdot|v_k)}{\mathcal{L}_{ij}(\cdot|v_k)} \frac{\partial \mathcal{L}_{ij}(\cdot|v_k)}{\partial \lambda} \quad (8.4.36)$$

$$\frac{\partial \ell_i}{\partial \gamma} = \frac{1}{\mathcal{L}_i(\cdot)} \sum_k p_k \sum_j \frac{\mathcal{L}_i(\cdot|v_k)}{\mathcal{L}_{ij}(\cdot|v_k)} \frac{\partial \mathcal{L}_{ij}(\cdot|v_k)}{\partial \gamma}. \quad (8.4.37)$$

$\mathcal{L}_{ij}(\cdot|v)$  tuletised on analoogilised nagu sõltumatute vaatluste korral. Lisaks tuleb  $p$  järgi gradiendi võtmisel arvestada, et  $\sum p_k = 1$ .

#### 8.4.4 Intervallandmed

Intervallandmetega on tegemist siis kui on vaadeldav ainult fakt et sündmus (seisundite vahetamine, tsenseerimine) toimus mingis kestuse(aja)vahemikus (näiteks kuu või nädala jooksul). Intervallmudel sobib ka siis kui ei soovi hasarti täpselt spetsifitseerida.

Mudel: olgu kestus jagatud  $T+1$  vahemikuks:  $[0, t_1), [t_1, t_2), \dots, [t_{T-1}, t_T), [t_T, \infty)$ . Iga indiviidi  $i$  kohta olgu vaadeldav et millises vahemikus ta algsest seisundist lahkus, või et millises vahemikust vaatlus on tsenseeritud. Eeldame MPH mudelit nagu 8.4.3. osas:

$$\vartheta(\tau|\mathbf{x}, v) = \lambda(\tau)e^{\beta' \mathbf{x}v}. \quad (8.4.38)$$

Tõenäosus, et isik jääb kogu intervalli  $n$  jooksul algseisundisse avaldub

$$S_n(\mathbf{x}, v) = \exp(-vz_n(\mathbf{x})) = \exp(-ve^{\beta' \mathbf{x}} \int_{\tau_{n-1}}^{\tau_n} \lambda(s) ds). \quad (8.4.39)$$

Nüüd võib defineerida  $\tilde{\lambda}_n$ :

$$e^{\tilde{\lambda}_n}(t_n - t_{n-1}) \equiv \int_{\tau_{n-1}}^{\tau_n} \lambda(s) ds, \quad (8.4.40)$$

kus  $e^{\tilde{\lambda}_s}$  on keskmine põhihasart vahemikus  $s$  ja  $\tilde{\lambda}$  on lihtsalt mudeli parameeter. Seega põhihasart on spetsifitseeritud mittepameetriselt. Vaatluse laiklihuud fikseeritud  $v$  korral avaldub nüüd

$$\mathcal{L}(n|\mathbf{x}, v) = (1 - e^{-vz_n(\mathbf{x})})^\delta \prod_{m=1}^{n-1} e^{-vz_m(\mathbf{x})}, \quad (8.4.41)$$

kus  $\delta = 1$  tähendab et vaatlus pole tsenseeritud. Kogu vaatluse log-laiklihuud on

$$\ell(n|\mathbf{x}) = \log \left( \sum_{k=1}^K p_k \mathcal{L}(n|\mathbf{x}, v_k) \right). \quad (8.4.42)$$

$v$ -spetsiifilise laiklihuudi gradient avaldub

$$\frac{\partial}{\partial \beta} \mathcal{L}(n|\mathbf{x}, v) = vE_g(n|\mathbf{x}, v) \frac{\partial}{\partial \beta} z_n(\mathbf{x}) - vS_g(n|\mathbf{x}, v) \sum_{m=1}^{n-1} \frac{\partial}{\partial \beta} z_m(\mathbf{x}) \quad (8.4.43)$$

$$\frac{\partial}{\partial \tilde{\lambda}_s} \mathcal{L}(n|\mathbf{x}, v) = vE_g(n|\mathbf{x}, v) \frac{\partial}{\partial \tilde{\lambda}_s} z_n(\mathbf{x}) - vS_g(n|\mathbf{x}, v) \sum_{m=1}^{n-1} \frac{\partial}{\partial \tilde{\lambda}_s} z_m(\mathbf{x}) \quad (8.4.44)$$

$$\frac{\partial}{\partial v} \mathcal{L}(n|\mathbf{x}, v) = E_g(n|\mathbf{x}, v) z_n(\mathbf{x}) - S_g(n|\mathbf{x}, v) \sum_{m=1}^{n-1} z_m(\mathbf{x}) \quad (8.4.45)$$

kus

$$E_g(n|\mathbf{x}, v) = \delta e^{-vz_n(\mathbf{x})} \prod_{m=1}^{n-1} e^{-vz_m(\mathbf{x})} \quad (8.4.46)$$

$$S_g(n|\mathbf{x}, v) = \mathcal{L}(n|\mathbf{x}, v) = \left(1 - e^{-v z_n(\mathbf{x})}\right)^\delta \prod_{m=1}^{n-1} e^{-v z_m(\mathbf{x})} \quad (8.4.47)$$

on gradiendi seisundist lahkumise ja seisundis kestmise spetsiifilised osad ja z gradient avaldub

$$\frac{\partial}{\partial \beta} z_n(\mathbf{x}) = z_n(\mathbf{x}) x_n \quad (8.4.48)$$

$$\frac{\partial}{\partial \tilde{\lambda}_s} z_n(\mathbf{x}) = z_n(\mathbf{x}) \mathbb{1}(s = n). \quad (8.4.49)$$

Kogulaiklihuudi gradient on

$$\frac{\partial}{\partial \beta} \ell(n|\mathbf{x}) = \frac{1}{\mathcal{L}(n|\mathbf{x})} \left( \sum_{k=1}^K p_k \frac{\partial}{\partial \beta} \mathcal{L}(n|\mathbf{x}, v) \right) \quad (8.4.50)$$

$$\frac{\partial}{\partial \tilde{\lambda}_s} \ell(n|\mathbf{x}) = \frac{1}{\mathcal{L}(n|\mathbf{x})} \left( \sum_{k=1}^K p_k \frac{\partial}{\partial \tilde{\lambda}_s} \mathcal{L}(n|\mathbf{x}, v) \right) \quad (8.4.51)$$

$$\frac{\partial}{\partial v_k} \ell(n|\mathbf{x}) = \frac{1}{\mathcal{L}(n|\mathbf{x})} p_k \frac{\partial}{\partial v_k} \mathcal{L}(n|\mathbf{x}, v) \quad (8.4.52)$$

$$\frac{\partial}{\partial p_k} \ell(n|\mathbf{x}) = \frac{\mathcal{L}(n|\mathbf{x}, v_k)}{\mathcal{L}(n|\mathbf{x})} \quad (8.4.53)$$

**Mitu lõppseisundit** Mitme lõppseisundiga mudelid esitatakse sageli *kompiiting risk* kujul. Too tähendab et kujutatakse ette et kõikidesse lõppseisunditesse viivad sõltumatud Markovi protsessid, realiseerub too mille aeg on kõige lühem. Vajadusel võib tsenseerimist kujutada ühena lõppseisunditest.

Kui kõik muutujad on vaadeldavad, ei erine mitme lõppseisundiga juhtub olukorrast kui modelleerida üksikuid lõppseisundeid sõltumatult, teised seisundid oleksid siis nagu tsenseeritud. Kui mudelis on mittevaadeldav heterogeensus, on pilt ainult veidi keerulisem.

Olgu  $M$  võimalikku lõppseisundit ja  $m \in \{1, \dots, M\}$  olgu lõppseisundi näitaja. Olgu seisundisse  $m$  ülemineku hasart, sõltuvalt vaadeldavatest ja mittevaadeldavatest parameetritest,  $\vartheta^m(\tau|x, v^m)$ . Mittevaadeldav heterogeensus olgu  $M$ -mõõtmelise diskreetse jaotusega:  $v^m \in \{v_1^m, \dots, v_{K^m}^m\}$  kusjuures  $v = (v_{k^1}^1, \dots, v_{k^M}^M)$  esineb tõenäosusega  $p_{k^1 \dots k^M}$ . Eeldame et erinevate spellide jooksul on  $v$  konstantne,  $x$  aga võib muutuda.

Inimese  $i$  spelli  $j$ , osa laiklihuudi funktsioonis siirde  $m$  järgi on nüüd:

$$\mathcal{L}_{ij}^m(\cdot|v^m) = \left[ \vartheta^m(\tau_{ij}|x_{ij}) \right]^{\delta_{ij}^m} S^m(\tau_{ij}|x_{ij}, v^m). \quad (8.4.54)$$

Siirdega seotud liige  $\left[ \vartheta^m(\tau_{ij}|x_{ij}) \right]^{\delta_{ij}^m}$  ja püsimisega seotud liige  $S^m(\tau_{ij}|x_{ij}, v^m)$  avalduvad nii nagu ühe spelli ja ühe lõppseisundi puhul (vaata näiteks 8.4.3 või 8.4.4). Siin  $\tau_{ij}$  on spelli vaadeldud kestus,  $\delta_{ij}^m$  on mitte-tsenseerituse indikaator  $m$ -seisundi mõttes (= 1 kui läks seisundisse  $m$  ja 0 kui es lähe) ja  $x_{ij}$  on inimese  $i$  vaadeldavad isikutunnused spelli  $j$  ajal. Spellide kogulaiklihuud on:

$$\mathcal{L}_{ij}(\cdot|x_{ij}, v) = \prod_{m=1}^M \mathcal{L}_{ij}^m(\cdot|x_{ij}, v^m). \quad (8.4.55)$$

Kompiiting risk mudel eeldab et siirded on sõltumatud (kui kontrollida  $x$  ja  $v$  suhtes), seega siis laiklihuudi korutus siirete kaupa.

Olgu inimese  $i$  kohta  $N_i$  vaadeldud spelli. Inimese  $i$  osa laiklihuudis avaldub siis

$$\mathcal{L}_i(\cdot|x, v) = \prod_{j=1}^{N_i} \mathcal{L}_{ij}(\cdot|x, v). \quad (8.4.56)$$

ja vaadeldav laiklihuud avaldub:

$$\mathcal{L}_i(\cdot|x) = \sum_{k^1=1}^{K^1} \dots \sum_{k^M=1}^{K^M} p_{k^1 \dots k^M} \mathcal{L}_i(\cdot|x, v). \quad (8.4.57)$$

Laiklihuudi gradient avaldub:

$$\frac{\partial \mathcal{L}_i(\cdot|x)}{\partial \lambda^m} = \sum_{k^1=1}^{K^1} \dots \sum_{k^M=1}^{K^M} p_{k^1 \dots k^M} \mathcal{L}_i(\cdot|x, v) \sum_{j=1}^{N_i} \left[ \frac{1}{\mathcal{L}_{ij}^m(\cdot|x v_{k^m}^m)} \frac{\partial \mathcal{L}_{ij}^m(\cdot|x v_{k^m}^m)}{\partial \lambda^m} \right] \quad (8.4.58)$$

$$\frac{\partial \mathcal{L}_i(\cdot|x)}{\partial \gamma^m} = \sum_{k^1=1}^{K^1} \dots \sum_{k^M=1}^{K^M} p_{k^1 \dots k^M} \mathcal{L}_i(\cdot|x, v) \sum_{j=1}^{N_i} \left[ \frac{1}{\mathcal{L}_{ij}^m(\cdot|x v_{k^m}^m)} \frac{\partial \mathcal{L}_{ij}^m(\cdot|x v_{k^m}^m)}{\partial \gamma^m} \right] \quad (8.4.59)$$

$$\frac{\partial \mathcal{L}_i(\cdot|x)}{\partial v_k^m} = \sum_{k^1=1}^{K^1} \dots \sum_{k^{m-1}=1}^{K^{m-1}} \sum_{k^{m+1}=1}^{K^{m+1}} \dots \sum_{k^M=1}^{K^M} p_{k^1 \dots k^{m-1} k^{m+1} \dots k^M} \cdot$$

$$\cdot \mathcal{L}_i(\cdot|\mathbf{x}, v) \sum_{j=1}^{N_i} \left[ \frac{1}{\mathcal{L}_{ij}^m(\cdot|\mathbf{x}v_k^m)} \frac{\partial \mathcal{L}_{ij}^m(\cdot|\mathbf{x}v_k^m)}{\partial \gamma^m} \right] \quad (8.4.60)$$

$$\frac{\partial \mathcal{L}_i(\cdot|\mathbf{x})}{\partial p_{k^1 \dots k^M}} = \mathcal{L}_i(\cdot|\mathbf{x}, (v_{k^1}^1, \dots, v_{k^M}^M)) \quad (8.4.61)$$

$\mathcal{L}_{ij}^m(\cdot|\mathbf{x}, v)$  tuletised on nii nagu sõltumatute vaatluste korral.



#### 8.4.5 Parametriseerimine

**Logistic transition probability in discrete-time** In discrete time, it is useful to parameterise the destination-specific exit probabilities logistically as

$$p^m \equiv \Pr(\text{exit to destination } m) = \frac{e^{\lambda^m}}{1 + \sum_k e^{\lambda^k}}. \quad (8.4.62)$$

The probabilities are guaranteed to be positive and their sum to be less than one while  $\lambda$ -s may be unbounded. The components of the corresponding gradient transformation matrix are

$$\frac{\partial}{\partial \lambda^k} p^m = \mathbb{1}(k = m) p_m - p_k p_m \quad (8.4.63)$$

and the inverse transformation

$$l^m = \log \frac{p^m}{1 - \sum_k p^k}. \quad (8.4.64)$$

**Transition probability in continuous time** A good choice for parametrising the time-dependent part of the hazard is

$$\lambda = e^{\tilde{\lambda}} \quad (8.4.65)$$

The  $\lambda$  is now guaranteed to be positive.

**Diskreetne mittevaaeldav heterogeensus** Diskreetne mittevaaeldav heterogeensus, üks lõppseisund:  $v \in \{v_1, v_2, \dots, v_K\}$  ja vastavad tõenäosused  $p_1, p_2, \dots, p_K$ .

Kui laiklihuudi maksimeerimisel kasutada vastavaid parameetreid  $\tilde{v}$  ning  $\tilde{p}$  ( $\tilde{N}$  parameetrit) kuid laiklihuudi arvutamiseks nad normaalkujule ( $N$  parameetrit) teisendada, siis peab arvestama, et vastavad gradiendi komponendid teisenevad:

$$\begin{pmatrix} \frac{\partial}{\partial \tilde{v}} \ell_i \\ \frac{\partial}{\partial \tilde{p}} \ell_i \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial \tilde{v}} v & \frac{\partial}{\partial \tilde{v}} p \\ \frac{\partial}{\partial \tilde{p}} v & \frac{\partial}{\partial \tilde{p}} p \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial v} \ell_i \\ \frac{\partial}{\partial p} \ell_i \end{pmatrix} \equiv \mathbf{C} \begin{pmatrix} \frac{\partial}{\partial v} \ell_i \\ \frac{\partial}{\partial p} \ell_i \end{pmatrix} \quad (8.4.66)$$

Maatriksi  $\mathbf{C}$  ridade arv vastab algkuju parameetrite arvule  $\tilde{N}$  ning veergude arv normaalkuju parameetrite arvule  $N$  (s.h. lineaarselt sõltuvad  $p_H$  ja  $v_H$ ). Kovarjatsjoonimaatriksi  $p$  ning  $v$  sisaldav osa on vastavalt:

$$\Sigma = \mathbf{C}' \tilde{\Sigma} \mathbf{C} \quad (8.4.67)$$

**Heterogeensus** Tõenäosused on võimalik parametriseerida kui

$$p_k = \frac{e^{\tilde{p}_k}}{\sum_{k=1}^K e^{\tilde{p}_k}} \quad \text{ja} \quad \sum_{k=1}^K \tilde{p}_k = 0, \quad (8.4.68)$$

kus  $K$  on jaotuse toetuspunktide arv.

**Keskvärtuse normeerimine**  $v$  keskvärtuse võib normeerida üheks:

$$v_k = e^{\tilde{v}_k} \quad \text{ja} \quad \sum_{k=1}^K p_k v_k = 1. \quad (8.4.69)$$

Vastav pöördteisendus tõenäosuste tarvis on

$$\tilde{p}_k = \log p_k - \frac{\sum_{l=1}^K \log p_l}{K}. \quad (8.4.70)$$

$C$  komponendid on:

$$\frac{\partial}{\partial \tilde{v}_l} v_k = \begin{cases} v_k & \text{kui } k = l < K \\ -v_l \frac{p_l}{p_K} & \text{kui } k = K \\ 0 & \text{muudel juhtudel} \end{cases} \quad (8.4.71)$$

$$\frac{\partial}{\partial \tilde{v}_l} p_k = 0 \quad (8.4.72)$$

$$\frac{\partial}{\partial \tilde{p}_l} v_k = \begin{cases} \frac{1}{p_K} [(1 - p_K) + p_K v_K + (1 - p_l) v_l] & \text{kui } k = K \\ 0 & \text{muudel juhtudel} \end{cases} \quad (8.4.73)$$

$$\frac{\partial p_k}{\partial \tilde{p}_l} = -p_k(p_l - p_K) + \mathbb{1}(k = l)p_K - \mathbb{1}(K = k)p_K \quad (8.4.74)$$

$v$  komponendi normeerimine Defineerime

$$v_K = 1. \quad (8.4.75)$$

Näib et nii on intervallandmete juures tulemused mõnevõrra stabiilsemad, samas on põhihasarti raskem tõlgendada.

$C$  komponendid on

$$\frac{\partial}{\partial \tilde{v}_l} v_k = \begin{cases} v_k & \text{kui } k = l < K \\ 0 & \text{muudel juhtudel} \end{cases} \quad (8.4.76)$$

$$\frac{\partial}{\partial \tilde{v}_l} p_k = 0 \quad (8.4.77)$$

$$\frac{\partial}{\partial \tilde{p}_l} v_k = 0 \quad (8.4.78)$$

$$\frac{\partial p_k}{\partial \tilde{p}_l} = -\frac{e^{\tilde{p}_k} (e^{\tilde{p}_l} - e^{\tilde{p}_H})}{\left(\sum_{i=1}^H e^{\tilde{p}_i}\right)^2} + \mathbb{1}(k = l) \frac{e^{\tilde{p}_k}}{\sum_{i=1}^H e^{\tilde{p}_i}} + \mathbb{1}(K = k) \frac{e^{\tilde{p}_K}}{\sum_{i=1}^K e^{\tilde{p}_i}} \quad (8.4.79)$$

$M$  lõppseisundit (8.4.66) asemel võib nüüd kirjutada:

$$\begin{pmatrix} \frac{\partial}{\partial \tilde{v}^1} \ell_i \\ \frac{\partial}{\partial \tilde{v}^2} \ell_i \\ \dots \\ \frac{\partial}{\partial \tilde{v}^M} \ell_i \\ \frac{\partial}{\partial \tilde{p}} \ell_i \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial \tilde{v}^1} v^1 & \frac{\partial}{\partial \tilde{v}^1} v^2 & \dots & \frac{\partial}{\partial \tilde{v}^1} v^M & \frac{\partial}{\partial \tilde{v}^1} p \\ \frac{\partial}{\partial \tilde{v}^2} v^1 & \frac{\partial}{\partial \tilde{v}^2} v^2 & \dots & \frac{\partial}{\partial \tilde{v}^2} v^M & \frac{\partial}{\partial \tilde{v}^2} p \\ \dots & \dots & \ddots & \dots & \dots \\ \frac{\partial}{\partial \tilde{v}^M} v^1 & \frac{\partial}{\partial \tilde{v}^M} v^2 & \dots & \frac{\partial}{\partial \tilde{v}^M} v^M & \frac{\partial}{\partial \tilde{v}^M} p \\ \frac{\partial}{\partial \tilde{p}} v^1 & \frac{\partial}{\partial \tilde{p}} v^2 & \dots & \frac{\partial}{\partial \tilde{p}} v^M & \frac{\partial}{\partial \tilde{p}} p \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial v^1} \ell_i \\ \frac{\partial}{\partial v^2} \ell_i \\ \dots \\ \frac{\partial}{\partial v^M} \ell_i \\ \frac{\partial}{\partial p} \ell_i \end{pmatrix} \\
\equiv \mathbf{C} \begin{pmatrix} \frac{\partial}{\partial v^1} \ell_i \\ \frac{\partial}{\partial v^2} \ell_i \\ \dots \\ \frac{\partial}{\partial v^M} \ell_i \\ \frac{\partial}{\partial p} \ell_i \end{pmatrix}, \quad (8.4.80)$$

kus  $\mathbf{C}$  on  $\tilde{N} \times N$  maatriks.

Olgu  $p_{Dk}^m$  suuna  $m$  spetsiifiline tõenäosus, s.t. millise tõenäosusega esineb väärtus  $v_k^m$ . Tuletised võib nüüd avaldada kui

$$\frac{\partial}{\partial \tilde{p}} v^m = \frac{\partial}{\partial \tilde{p}} p \frac{\partial}{\partial p} p_D^m \frac{\partial}{\partial p_D^m} v^m \quad (8.4.81)$$

## 9 Algorithms

### 9.1 Arrays

**Array indexing** Let  $M$  be a  $N$ -dimensional column-major zero-based array, and  $m[i_1, i_2, \dots, i_N]$  its element where  $i_j \in \{0, \dots, K_j - 1\}$  and  $K_j$  is the size of dimension  $j$ . Vectorized index for element  $m[i_1, i_2, \dots, i_N]$  is

$$j = i_1 + K_1 i_2 + K_2 i_3 + \dots + K_{N-1} i_N. \quad (9.1.1)$$

The array indices can be recovered from the vector index  $j$  as:

$$\begin{aligned} i_1 &= j \bmod K_1 \\ i_2 &= [j \bmod (K_1 K_2)] // K_1 \\ i_3 &= [j \bmod (K_1 K_2 K_3)] // (K_1 K_2) \\ i_4 &= [j \bmod (K_1 K_2 K_3 K_4)] // (K_1 K_2 K_3) \\ &\dots \end{aligned}$$

where  $//$  is the integer division.

## Index

beta function, [9](#)

digamma function, [8](#), [14](#)

Frobenius norm, [28](#)

gamma distribution, [40](#)

inverse matrix, [27](#)

Lipschitz condition, [11](#)

Lipschitz continuity, *see* Lipschitz condition

log-normal, [42](#)

logarithm, [10](#)

logistic distribution, [43](#)

Moore-Penrose pseudoinverse, [27](#)

softmax, [9](#)

subderivative, [13](#)

supporting hyperplane, [6](#)

Toeplitz matrix, [27](#)

## References

- Calzolari, G. and Fiorentini, G. (1993) Alternative covariance estimators of the standard Tobit model, **42**, 5–13.
- Chiang, A. C. (1984) *Fundamental Methods of Mathematical Economics*, McGraw-Hill, Singapore, third edn.
- Lin, J. (1991) Divergence measures based on the shannon entropy, *Information Theory, IEEE Transactions on*, **37**, 145–151.
- Miller, R. E. (1979) *Dynamic Optimization and Economic Applications*, McGraw-Hill, New York.
- Murphy, K. P. (2012) *Machine Learning: A Probabilistic Perspective*, MIT Press, Cambridge, MA.
- Newey, W. K. (1991) Uniform convergence in probability and stochastic equicontinuity, *Econometrica*, **59**, 1161–1167.