Mathematical Formulas and Other Valuable Knowledge

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

Ott Toomet

Seattle, August 29, 2020

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 refence. 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 pieces of text from that time, in particlar 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, August 29, 2020

Contents

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:

$$\begin{pmatrix} \mathrm{d}u \\ \mathrm{d}v \end{pmatrix} = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix} \begin{pmatrix} \mathrm{d}x \\ \mathrm{d}y \end{pmatrix}, \tag{1.1.1}$$

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

$$\mathrm{d}S = \mathrm{d}x\mathrm{d}y = \frac{\mathrm{d}u\mathrm{d}v}{|\det \mathcal{J}|}, \tag{1.1.2}$$

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

1.2 Kolmnurk

1.2.1 Seos kolmnurga külgede ja nurkade vahel

$$a^2 + b^2 - 2ab\cos\gamma = c^2. \tag{1.2.1}$$

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

$$a^2 + b^2 = c^2. \tag{1.2.2}$$

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

$$S = \frac{1}{2}|X_A \cdot Y_B - Y_A \cdot X_B| = \frac{1}{2}\mathrm{abs}\begin{vmatrix} X_A & Y_A \\ X_B & X_B \end{vmatrix}. \tag{1.2.3}$$

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

1.3 Trigonometrics

Trigonometric Identities

$$\begin{aligned} \sin(2\alpha) &= 2\cos\alpha\, \sin\alpha \\ \cos(2\alpha) &= \cos^2\alpha - \sin^2\alpha \\ \sin(\alpha+\beta) &= \cos\alpha\, \sin\beta + \sin\alpha\, \cos\beta \\ \cos(\alpha+\beta) &= \cos\alpha\, \cos\beta - \sin\alpha\, \sin\beta \end{aligned}$$

1.4 Joone kõverus

Parameetriliselt antud joone kõverus:

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

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 General

Bijective function (also bijection) is an one-to-one relationship between elements of two sets.

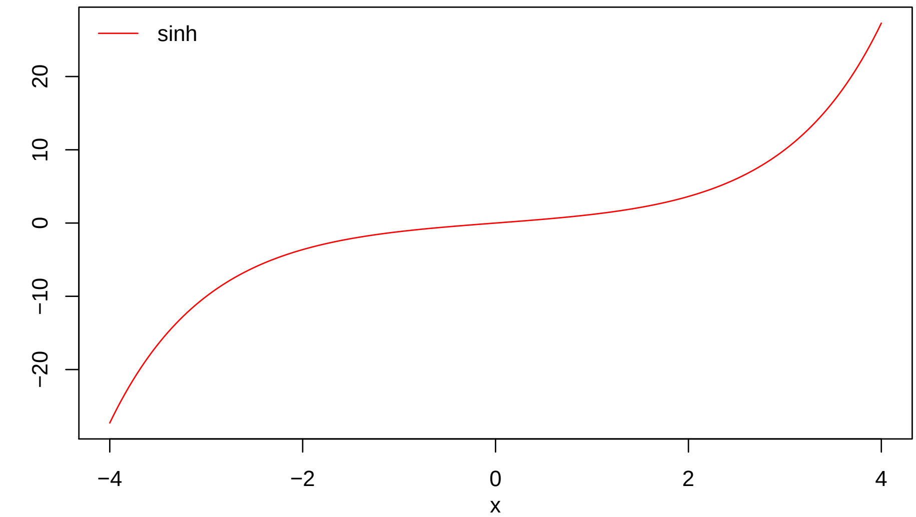
2.2 Elementary functions

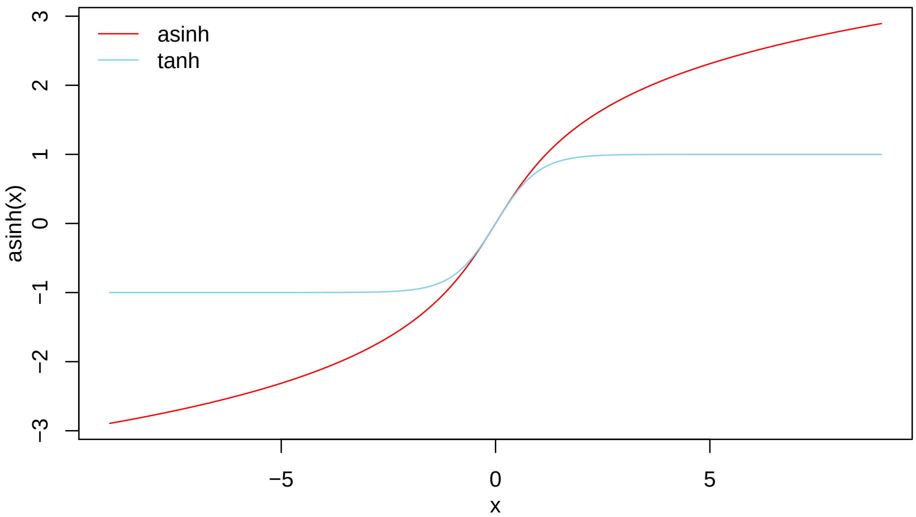
Hyperbolic functions

$$\sinh x = \frac{e^x - e^{-x}}{2} \tag{2.2.1}$$

$$\operatorname{arcsinh} x = \log\left(x + \sqrt{x^2 + 1}\right) \tag{2.2.2}$$

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^{2x} - 1}{e^{2x} + 1} \tag{2.2.3}$$





A closely related to tanh is *hard tanh*:

$$\text{hard}\,\tanh(x) = \begin{cases} -1 & \text{if} \quad x < -1 \\ x & \text{if} \quad -1 \leq x \leq 1 \\ 1 & \text{if} \quad x > 1 \end{cases}$$

Derivatives

$$\frac{\partial}{\partial x} \tanh(x) = 4 \frac{e^{2x}}{(e^{2x}+1)^2} = 1 - \tanh^2(x) \tag{2.2.4}$$

2.3 Algebraic Functions

Gamma Function $\Gamma(\cdot)$

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt \tag{2.3.1}$$

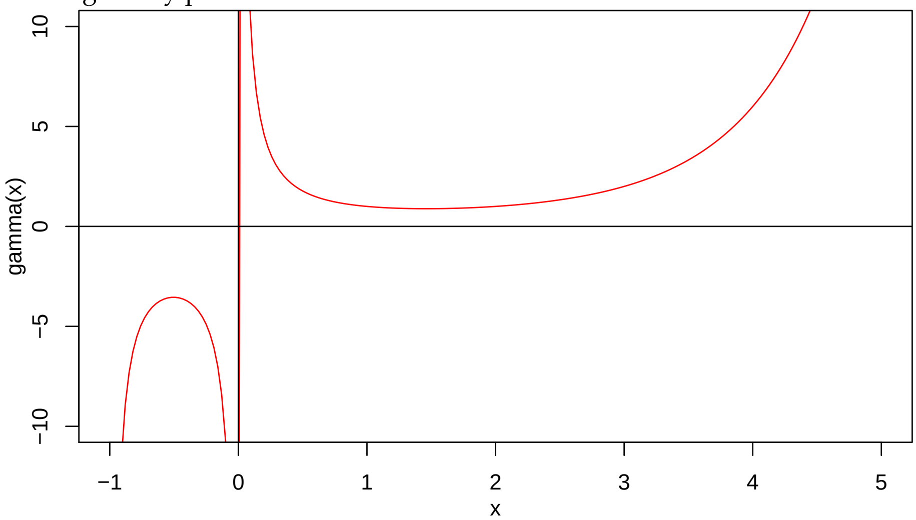
Properties:

$$\Gamma(1) = \Gamma(2) = 1 \tag{2.3.2}$$

$$\Gamma(n) = (n-1)! \tag{2.3.3}$$

$$\Gamma(n+1) = n\Gamma(n) \tag{2.3.4}$$

Proof: integrate by parts.



Digamma function

$$\psi(x) = \frac{\partial}{\partial x} \log \Gamma(x) = \frac{1}{\Gamma(x)} \int_0^\infty t^{x-1} e^{-t} \log t\, dt. \tag{2.3.5}$$

Properties:

$$\psi(\alpha+1) = \frac{1}{\alpha} + \psi(\alpha) \tag{2.3.6}$$

$$\psi(1) = -0.5772 \qquad\qquad \psi(2) = 0.4228 \tag{2.3.7}$$

Multivariate gamma function

$$\Gamma_D(x) = \pi^{\frac{1}{4}D(D-1)} \prod_{i=1}^{D} \Gamma\left(x + \frac{1-i}{2}\right) \tag{2.3.8}$$

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

Modified Bessel Functions Modified Bessel function of the first kind:

$$I_\alpha(x) = \sum_{k=0}^{\infty} \frac{1}{k! \cdot \Gamma(\alpha + k + 1)} \left(\frac{x}{2}\right)^{2k+\alpha} \tag{2.3.9}$$

Modified Bessel function of the second kind:

$$K_\alpha(x) = \frac{\pi}{2} \frac{I_{-\alpha}(x) - I_\alpha(x)}{\sin \alpha\pi} \tag{2.3.10}$$

Beta function $B(a, b)$

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} = \int_0^1 x^{p-1}(1-x)^{q-1}\,\mathrm{d}x, \tag{2.3.11}$$

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

2.4 Other functions

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

$$S(\boldsymbol{x})_i = \frac{\mathrm{e}^{x_i}}{\sum_j \mathrm{e}^{x_j}} \qquad i \in \{1, 2, \ldots, K\}. \tag{2.4.1}$$

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

Derivatives:

$$\frac{\partial}{\partial x_j} S(x)_i = \begin{cases} -S(x)_i \cdot S(x)_j & \text{if} \quad i \neq j \\ S(x)_i \cdot (1 - S(x)_j) & \text{if} \quad i = j \end{cases} \tag{2.4.2}$$

$$\frac{\partial}{\partial x_j} \log S(x)_i = \begin{cases} -S(x)_j & \text{if} \quad i \neq j \\ 1 - S(x)_j & \text{if} \quad i = j \end{cases} \tag{2.4.3}$$

Softplus function In 1-dimensional case

$$f(x) = \log(1 + \mathrm{e}^x). \tag{2.4.4}$$

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

2.5 Variable transformations

Logistic transformation Logistic transformation can be used to map $\mathbb{R} \to [0, 1]$:

$$y = \Lambda(x) = \frac{1}{1 + \mathrm{e}^{-x}} \qquad x = \log \frac{y}{1 - y} \tag{2.5.1}$$

3 Calculus

3.1 Logarithm

Properties

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

3.2 Limits

3.2.1 Limits in general

$\liminf$ **and** $\limsup$ Definition:

$$\liminf_{n\to\infty} \equiv \lim_{n\to\infty}\left(\inf_{m\geq n} x_m\right) \tag{3.2.1}$$

L'Hôpital's rule Let f and g be differentiable functions, except possibly at c, and $g'(x) \neq 0$ if $x \neq c$, then

$$\lim_{x\to c}\frac{f(x)}{g(x)} = \frac{f'(x)}{g'(x)} \tag{3.2.2}$$

if $\lim_{x\to c} f(x) = \lim_{x\to c} g(x) = 0$, or ∞.

3.2.2 Other limits

$$\lim_{x\to 0} x \cdot \log x = 0 \tag{3.2.3}$$

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

$$\lim_{x\to\infty}\frac{1}{x}\frac{\phi(x)}{1-\Phi(x)} = 1, \tag{3.2.4}$$

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

Proof. The fraction can be written as

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

The integral can be expressed as the Rieman limit

$$\int_x^\infty \phi(s)\mathrm{d}s = \lim_{\delta\to 0}\left[\phi(x)\delta + \phi(x+\delta)\delta + \phi(x+2\delta)\delta + \ldots\right].$$

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

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

and hence we may write the denominator in (3.2.4) as

$$x\phi(x)\lim_{\delta\to 0}\left[1 + \mathrm{e}^{-x\delta}\mathrm{e}^{-\delta^2/2} + \mathrm{e}^{-2x\delta}\mathrm{e}^{-4\delta^2/2} + \mathrm{e}^{-3x\delta}\mathrm{e}^{-9\delta^2/2} + \ldots\right]\delta.$$

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 $x\delta \to 0$. Accordingly, we let $\delta \to 0$ and $x \to \infty$ in such as way that $x\delta \to 0$. 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 $1 + e^{-x\delta} + e^{-2x\delta} + \ldots$ is less than $\varepsilon > 0$:

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

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 $x \to \infty$ and $\delta \to 0$, the exponent converges to 1. Hence, at the limit, the Rieman sum is solely determined by the geometric sequence, and we have denominator in (3.2.4) equal to $\phi(x)$. $\square$

3.2.3 Limits Involving Factorial

Stirling's formula

$$\log n! = n \log n - n + O(\log n) \tag{3.2.5}$$

Or a more precise version

$$\log n! \approx \frac{1}{2}\log(2\pi n) + n \log n - n \tag{3.2.6}$$

faktoriaalide jagatis

$$\lim_{n\to\infty} \frac{n!}{(n-m)!} = n^m \tag{3.2.7}$$

Märkus: siin on eeldatud, et $m \not\to \infty$. Jaga läbi, arvesta, et $n - 1 \approx n$. A special case:

$$\lim_{n\to\infty} \binom{n}{m} = \lim_{n\to\infty} \frac{n!}{(n-m)!m!} = \frac{n^m}{m!}. \tag{3.2.8}$$

3.3 Boundedness and Convergence

3.3.1 General Concepts

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

$$d(f_i(x), a) \le M \quad \forall i \in \mathcal{I}, x \in X \tag{3.3.1}$$

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

$$|f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)| \leq L \cdot d(\boldsymbol{x}_1, \boldsymbol{x}_2) \quad \forall \boldsymbol{x}_1, \boldsymbol{x}_2 \text{ in range of } f(\cdot) \tag{3.3.2}$$

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 $\mathcal{F} = \{f_i : X \to Y, i \in \mathcal{I}\}$ is equicontinuous at point x if for every $\epsilon > 0$ there exists $\delta(\epsilon, x)$ such that

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

Note: δ must not depend on i, but may depend on x, ϵ.

3.3.2 Tests

Cauchy root test Series $\sum_{n=1}^{\infty} a_n$ converges if

$$R = \limsup_{n \to \infty} \sqrt[n]{|a_n|} < 1. \tag{3.3.4}$$

If $R > 1$ the series diverges.

3.4 Differentiation

3.4.1 Concepts

Directional Derivative

$$f'(\boldsymbol{x};\boldsymbol{u}) = \lim_{h\to 0} \frac{f(\boldsymbol{x}+h\boldsymbol{u}) - f(\boldsymbol{x})}{h} \tag{3.4.1}$$

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

$$f(x) - f(x_0) \geq c(x - x_0) \tag{3.4.2}$$

for all x in I.

Jacobian Matrix

$$\mathcal{J} = \begin{bmatrix} \dfrac{\partial \mathbf{f}}{\partial x_1} & \cdots & \dfrac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \frac{\partial(f_1,\ldots,f_n)}{\partial(x_1,\ldots,x_n)} = \boldsymbol{D}\,\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{3.4.3}$$

Envelope theorem Olgu M defineeritud kui optimum funktsioonist f:

$$M(a) = \max_x f(x,a), \tag{3.4.4}$$

kus a on parameeter. Siis

$$\frac{\mathrm{d}M(a)}{\mathrm{d}a} = \left.\frac{\partial f(x^*,a)}{\partial a}\right|_{x^*=x(a)}. \tag{3.4.5}$$

3.4.2 Simple Derivatives

$$\frac{\partial}{\partial x} a^x = a^x \log a \tag{3.4.6}$$

$$\tan \phi' = \frac{1}{\cos^2 \phi} \tag{3.4.7}$$

$$\arcsin x' = \frac{1}{\sqrt{1-x^2}} \tag{3.4.8}$$

$$\arctan x' = \frac{1}{1+x^2} \tag{3.4.9}$$

$$\frac{\partial}{\partial x} \sqrt{1-x^2} = -\frac{x}{\sqrt{1-x^2}} = -\frac{x}{y} \tag{3.4.10}$$

$$\frac{\partial^2}{\partial x^2} \sqrt{1-x^2} = -\frac{1}{\sqrt{1-x^2}}\left(1+\frac{x^2}{1-x^2}\right) = -\frac{1}{y^3} \tag{3.4.11}$$

$$\frac{\partial}{\partial x} \sqrt{\frac{1-x^2}{1+x^2}} = -\frac{2}{\sqrt{(1-x^2)(1+x^2)}}\frac{x}{1+x^2} = -\frac{x}{y}\frac{1+y^2}{1+x^2} \tag{3.4.12}$$

3.4.3 Normal Density Related Derivatives

One-Dimensional Case

$$\frac{\mathrm{d}}{\mathrm{d}x}\Phi(-x) = \phi(x) \tag{3.4.13}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\phi\left(\frac{x-\mu}{\sigma}\right) = -\frac{1}{\sigma}\left(\frac{x-\mu}{\sigma}\right)\phi\left(\frac{x-\mu}{\sigma}\right) \tag{3.4.14}$$

$$\frac{\mathrm{d}}{\mathrm{d}\sigma}\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\left[\left(\frac{x-\mu}{\sigma}\right)^2 - 1\right]\phi\left(\frac{x-\mu}{\sigma}\right) \tag{3.4.15}$$

Multi-Dimensional Case Let $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ be the 2-dimensional variance-covariance matrix and $\phi(\cdot,\cdot)$ the 2-D normal density, defined in (5.8.6):

$$\frac{\partial}{\partial x_1}\phi(\boldsymbol{x},\Sigma) = \phi(\boldsymbol{x},\Sigma)\frac{\rho x_2 - x_1}{1-\rho^2} \tag{3.4.16}$$

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

$$= \phi(\boldsymbol{x},\Sigma)\left[\frac{\rho}{1-\rho^2} - \frac{\rho}{(1-\rho^2)^2}(x_1^2 - 2\rho x_1 x_2 + x_2^2) + \frac{x_1 x_2}{1-\rho^2}\right] \tag{3.4.17}$$

$$\frac{\partial^2}{\partial x_1 \partial x_2}\phi(\boldsymbol{x},\Sigma) = \phi(\boldsymbol{x},\Sigma)\left[\frac{(x_1-\rho x_2)(x_2-\rho x_1)}{(1-\rho^2)^2} + \frac{\rho}{1-\rho^2}\right] \tag{3.4.18}$$

3.4.4 Derivatives of Gamma Function

$$\Gamma'(x) = \psi(x)\Gamma(x) \tag{3.4.19}$$

$$\Gamma''(x) = \Gamma(x)\left[\psi'(x) + \psi^2(x)\right], \tag{3.4.20}$$

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

3.4.5 Differentiation of Sums

$$\frac{\partial}{\partial x_i}\sum_j x_j = 1 \tag{3.4.21}$$

$$\frac{\partial}{\partial x_i}\left(\sum_j x_j\right)^2 = 2\sum_j x_j \tag{3.4.22}$$

3.4.6 General Differentiation Rules

Pöördfunktsiooni tuletis

$$\frac{\mathrm{d}}{\mathrm{d}x}f^{-1}(x) = \left.\frac{1}{\frac{\mathrm{d}}{\mathrm{d}x}f(x)}\right|_{x=f^{-1}(y)} = \frac{1}{f'[f^{-1}(y)]} \tag{3.4.23}$$

Chain Rule Let $y = y(x)$ and $x = x(z)$ so $y = y(x(z)) \equiv f(z)$. The derivatives are:

$$\frac{\partial f(z)}{\partial z} = \frac{\partial y(x)}{\partial x} \cdot \frac{\partial x(z)}{\partial z} \equiv y'(x) \cdot x'(z) \tag{3.4.24}$$

$$\frac{\partial^2 y(x(z))}{\partial z^2} \equiv \frac{\partial^2 f(z)}{\partial z^2} = \frac{\partial^2 y(x)}{\partial x^2} \left(\frac{\partial x(z)}{\partial z}\right)^2 + \frac{\partial y(x)}{\partial x} \frac{\partial^2 x(z)}{\partial z^2} \tag{3.4.25}$$

See also the vector form in Section 4.8.2.

If $x(z)$ is a linear function, then the second term of $\frac{\partial^2 y(x(z))}{\partial b^2}$ will fall away and we have $\frac{\partial^2 f(z)}{\partial z^2} = \frac{\partial^2 y(x)}{\partial x^2} \left(\frac{\partial x(z)}{\partial z}\right)^2$.

Implicit Function Differentiation Define $y = g(x)$ implicitly through $f(x, y) = 0$. The first differential

$$\mathrm{d}f(x, y) = f_x(x, y)\,\mathrm{d}x + f_y(x, y)\,\mathrm{d}y = 0 \quad \Rightarrow \quad g'(y) = \frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{f_x(x, y)}{f_y(x, y)} \tag{3.4.26}$$

The second differential

$$\begin{aligned} &\mathrm{d}(f_x + f_y y') = \\ &\quad f_{xx}\,\mathrm{d}x + f_{yx} y'\,\mathrm{d}x + f_y y'_x\,\mathrm{d}x + f_{yy}(y')^2\,\mathrm{d}y + f_{xy}\,\mathrm{d}y + f_y y'_y\,\mathrm{d}y \\ &\qquad = 0 \end{aligned} \tag{3.4.27}$$

Note that y'_y, the partial derivative of y' w.r.t x, is 0. Now

$$y'' = y'_x = -\frac{1}{f_y}\left(f_{xx} + 2 f_{xy} y' + f_{yy} y'^2\right) \tag{3.4.28}$$

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

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

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

3.5.2 Integrals related to probability distributions

Normal distribution Let

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

and define

$$\int \phi(x)\,dx \equiv \Phi(x) \tag{3.5.4}$$

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

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

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

$$\int x\phi(x)\,dx = -\phi(x) \tag{3.5.7}$$

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

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

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

$$\int (x-\mu)^2\phi(x-\mu)\,dx = \Phi(x-\mu) - (x-\mu)\phi(x-\mu) \tag{3.5.11}$$

$$\int x^2\phi\left(\frac{x-\mu}{\sigma}\right) dx = \sigma(\mu^2+\sigma^2)\Phi\left(\frac{x-\mu}{\sigma}\right) - \sigma^2(x+\mu)\phi\left(\frac{x-\mu}{\sigma}\right) \tag{3.5.12}$$

$$\int_a^b x^2\phi(x)\,dx = a\phi(a) - b\phi(b) + \Phi(b) - \Phi(a) \tag{3.5.13}$$

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

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

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

$$\int_s^t \phi(u)\log\phi(u-\mu)\,du = \frac{1}{2}\left[\Phi(s)-\Phi(t)\right]\left(1+\log 2\pi+\mu^2\right) + \\ + \frac{1}{2}\left[t\phi(t)-s\phi(s)\right] + \mu\left[\phi(s)-\phi(t)\right] \tag{3.5.17}$$

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

$$\int \phi(x)\Phi(x)\,\mathrm{d}x = \frac{1}{2}\Phi^2(x) \tag{3.5.19}$$

$$\int x\phi(x)\Phi(x)\,\mathrm{d}x = -\phi(x)\Phi(x) + \frac{1}{2\sqrt{\pi}}\Phi(\sqrt{2}\,x) \tag{3.5.20}$$

$$\int x^2\phi(x)\Phi(x)\,\mathrm{d}x = \frac{1}{2}\Phi^2(x) - x\phi(x)\Phi(x) - \frac{1}{2}\phi^2(x) \tag{3.5.21}$$

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

$$\int_a^\infty xf(x)\,\mathrm{d}x = \int_a^b \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2\right]\mathrm{d}x =$$

$$= \frac{1-\Phi\left(\frac{\log a-\mu-\sigma^2}{\sigma}\right)}{1-\Phi\left(\frac{\log a-\mu}{\sigma}\right)}\mathrm{e}^{\mu+\frac{1}{2}\sigma^2} \tag{3.5.22}$$

$$\int_a^b xf(x)\,\mathrm{d}x = \int_a^b \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2\right]\mathrm{d}x =$$

$$= \mathrm{e}^{\mu+\frac{1}{2}\sigma^2}\left[-\Phi\left(\frac{\log a-\mu-\sigma^2}{\sigma}\right)+\Phi\left(\frac{\log b-\mu-\sigma^2}{\sigma}\right)\right] \tag{3.5.23}$$

3.5.3 Other integrals

$$\int_0^t \mathrm{e}^{-rT}\,\mathrm{d}T = \frac{1}{r}\left[1-\mathrm{e}^{-rt}\right] \tag{3.5.24}$$

$$\int \frac{\mathrm{d}x}{x(1-x)} = \log x - \log(1-x) \tag{3.5.25}$$

$$\int_{t_0}^T \frac{\mathrm{d}T}{T(1-T)} = \log\left[\frac{T}{1-T}\frac{1-t_0}{t_0}\right] \tag{3.5.26}$$

$$\int \frac{\mathrm{d}x}{(p+q^{ax})^2}\,\mathrm{d}x = \frac{x}{p^2} + \frac{1}{ap(p+q\mathrm{e}^{ax})} - \frac{1}{ap^2}\log(p+q\mathrm{e}^{ax}) \tag{3.5.27}$$

$$\int \log x\,\mathrm{d}x = x\log x - x \tag{3.5.28}$$

$$\int \mathrm{e}^{-ax}\mathrm{d}x = \frac{1}{a}\left[1-\mathrm{e}^{-ax}\right] \tag{3.5.29}$$

$$\int x\mathrm{e}^{x}\mathrm{d}x = x\mathrm{e}^{x} - \mathrm{e}^{x} \tag{3.5.30}$$

$$\int x\mathrm{e}^{-x}\mathrm{d}x = -x\mathrm{e}^{-x} - \mathrm{e}^{-x} \tag{3.5.31}$$

$$\int x\mathrm{e}^{\alpha x}\,\mathrm{d}x = \frac{x}{\alpha}\mathrm{e}^{\alpha x} - \frac{1}{\alpha^2}\mathrm{e}^{\alpha x} \tag{3.5.32}$$

$$\int_a^b x\mathrm{e}^{\alpha x}\,\mathrm{d}x = \frac{1}{\alpha}\left[\mathrm{e}^{\alpha b}\left(b-\frac{1}{\alpha}\right)-\mathrm{e}^{\alpha a}\left(a-\frac{1}{\alpha}\right)\right] \tag{3.5.33}$$

$$\int x^2 e^{\alpha a}\,dx = \frac{1}{\alpha}x^2 e^{\alpha x} - \frac{2}{\alpha}\int x e^{\alpha x}\,dx \tag{3.5.34}$$

$$\int_a^b x^2 e^{\alpha a}\,dx = \frac{1}{\alpha}\left[e^{\alpha b}\left(b^2 - \frac{2b}{\alpha} + \frac{2}{\alpha^2}\right) - e^{\alpha a}\left(a^2 - \frac{2a}{\alpha} + \frac{2}{\alpha^2}\right)\right] \tag{3.5.35}$$

$$\int_0^\infty x^\alpha e^{-\beta x} dx = \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \tag{3.5.36}$$

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

$$\int_c^b \left[f(x)\int_c^x w(y)\,dy\right] dx = \int_c^b \bar{F}(x) w(x)\,dx \tag{3.5.37}$$

Euleri konstant

$$\int_0^\infty e^{-z}\log z\,dz = c \approx -0{,}5772 \tag{3.5.38}$$

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

$$\begin{aligned}\int_V f(x_1,\dots,x_N)dx_1\dots dx_N &= \\ &= \int_V f[y_1(x_1,\dots,x_N),\dots,y_N(x_1,\dots,x_N)]\frac{dy_1\dots dy_N}{|\mathcal{J}|} = \\ &= \int_V g(y_1,\dots,y_N)\frac{dy_1\dots dy_N}{|\mathcal{J}|},\end{aligned} \tag{3.5.39}$$

kus

$$|\mathcal{J}| = \left|\frac{\partial(y_1,\dots,y_N)}{\partial(x_1,\dots,x_N)}\right| \tag{3.5.40}$$

on koordinaatteisenduse jakobjaani absoluutväärtus.

3.5.5 Laplace'i teisendus

indexLaplace transform|textbf

Laplace'i teisendus juhusliku muutuja X jaotusfunktsioonist on

$$L_f(s) = \mathbb{E}\,e^{-sX} = \int e^{-sx} dF_X(x). \tag{3.5.41}$$

Laplace'i teisendus on sama mis momendifunktsioon.

3.5.6 Numbriline integreerimine

Monte-Carlo integraal Olgu vaja leida

$$I = \int_a^b f(x)\,dx = (b-a)\int_a^b f(x)\frac{1}{b-a}\,dx = (b-a)\,\mathbb{E}[f(X)], \tag{3.5.42}$$

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

$$\hat{I}_N = (b-a)\frac{1}{N}\sum_{i=1}^{N} f(x_i) \tag{3.5.43}$$

$$\widehat{\operatorname{Var}\hat{I}_N} = \frac{(b-a)^2}{N}\frac{1}{N}\sum_{i=1}^{N}\left[f(x_i) - \frac{1}{N}\hat{I}_N\right]^2 \tag{3.5.44}$$

3.6 Differential Equations

Let c be a constant.

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

$$\begin{aligned} y(x)' + P(x)y &= Q(x) \\ y(x) &= \mathrm{e}^{-\int P(x)\,\mathrm{d}x} \int Q(x)\mathrm{e}^{\int P(x)\,\mathrm{d}x}\,\mathrm{d}x + c\mathrm{e}^{-\int P(x)\,\mathrm{d}x} \end{aligned} \tag{3.6.2}$$

3.7 Optimization

3.7.1 Second-order maximum/minimum conditions for constrained optimisation

The problem is

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

Corresponding Lagrangian is

$$Z = f(x_1, x_2, \ldots, x_n) - \lambda g(x_1, x_2, \ldots, x_n). \tag{3.7.2}$$

Corresponding *bordered Hessian* is:

$$\left|\bar{\mathsf{H}}\right| = \begin{vmatrix} 0 & g_1 & g_2 & \cdots & g_n \\ g_1 & Z_{11} & Z_{12} & \cdots & Z_{1n} \\ g_2 & Z_{21} & Z_{22} & \cdots & Z_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_n & Z_{n1} & Z_{n2} & \cdots & Z_{nn} \end{vmatrix} \tag{3.7.3}$$

and successive *principal minors* are:

$$\left|\bar{\mathsf{H}}_2\right| = \begin{vmatrix} 0 & g_1 & g_2 \\ g_1 & Z_{11} & Z_{12} \\ g_2 & Z_{21} & Z_{22} \end{vmatrix} \qquad \left|\bar{\mathsf{H}}_3\right| = \begin{vmatrix} 0 & g_1 & g_2 & g_3 \\ g_1 & Z_{11} & Z_{12} & Z_{13} \\ g_2 & Z_{21} & Z_{22} & Z_{23} \\ g_3 & Z_{31} & Z_{32} & Z_{33} \end{vmatrix} \qquad \ldots \qquad \left|\bar{\mathsf{H}}_n\right| = \left|\bar{\mathsf{H}}\right| \tag{3.7.4}$$

The second derivative $\mathrm{d}^2 z$ is positive definite iff

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

and negavitve definite iff

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

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

Proof: Chiang (1984).

3.7.2 Optimal control (dynamic optimisation)

The problem is:

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

Corresponding *Hamiltonian* is

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

The first order conditions for optimum are

1. $\max_u H(t, y, u, \lambda) \quad \forall t \in [0, T]$ or, less generally, $\frac{\partial H}{\partial t} = 0$ (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 λ).
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-dimesional parameter vector. Let the initial value of the parameter be $\boldsymbol{\vartheta}_0$. From Taylor's approximation:

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

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):

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

The algorith 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 equlity for approximating the Hessian:

$$\mathbb{E}\left[\frac{\partial^2 l(\vartheta)}{\partial\vartheta\partial\vartheta'}\right]_{\vartheta=\vartheta_0} = -\,\mathbb{E}\left[\left.\frac{\partial l(\vartheta)}{\partial\vartheta'}\right|_{\vartheta=\vartheta_0} \left.\frac{\partial l(\vartheta)}{\partial\vartheta}\right|_{\vartheta=\vartheta_0}\right] \tag{3.7.9}$$

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 $B \nsubseteq A$.

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

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

4.2.2 Basic Operations

De Morgan's Laws

$$\Big(\bigcap_{A \in F} A\Big)^C = \bigcup_{A \in F} A^C \tag{4.2.2}$$

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

Other

$$A \setminus B = A \cap B^C \tag{4.2.4}$$

$$\Big(\bigcup_i A_i\Big) \setminus (\bigcup_i B_i) \subseteq \bigcup_i (A_i \setminus B_i) \tag{4.2.5}$$

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: $\cup_i^n E_i \in S \quad E_i \in S, i = 1, 2, \ldots$
4. S is closed under countable intersection: $\cap_i^n E_i \in S \quad E_i \in S, i = 1, 2, \ldots$. This follows from 2,3 and De Morgan's laws.

4.3 Simple Algebra

Binomial Theorem

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

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

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

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

Geomeetrilise jada summa

$$S = 1 + q + q^2 + q^3 + ... = \frac{1}{1-q}. \tag{4.3.4}$$

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

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

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

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

Eksponent piirväärtusena

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

Tõestus: arenda Newtoni binoomvalemiga ritta, arvesta (3.2.7).

4.4 Taylori rida

Taylori rida Iga funktsiooni võib punkti x_0 ümbruses esitada astmereana:

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2} f''(x_0)(x-x_0)^2 + ... = \\ = \sum_{i=0}^{\infty} f^{(i)}(x_0) \frac{(x-x_0)^i}{i!}. \tag{4.4.1}$$

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:

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

Eksponendi astmerida

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

Tõestus: arenda e^x Taylori ritta.

Eksponendi piirväärtus

$$\lim_{x \to 0} e^x = 1 + x. \tag{4.4.4}$$

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

Logaritmi astmerida

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

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

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

4.5 Summade astmed

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

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

$$\left(\sum_{i=1}^{N} x_i\right)^4 = \sum_{i=1}^{N} x_i^4 + 4\sum_{\substack{i;j=1\\ i\neq j}}^{N} x_i x_j^3 + 3\sum_{\substack{i;j=1\\ i\neq j}}^{N} x_i^2 x_j^2 + 6\sum_{\substack{i;j;k=1\\ i\neq j;j\neq k;k\neq i}}^{N} x_i x_j x_k^2 +$$

$$+ \sum_{\substack{i;j;k;l=1\\ i;j;k;l\neq}}^{N} x_i x_j x_k x_l \tag{4.5.3}$$

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

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

$$\mathbb{E}\left(\sum^{N} x_i\right)^2 = N\,\mathbb{E}\,X^2 + N(N-1)(\mathbb{E}\,X)^2 = N^2(\mathbb{E}\,X)^2 + N\mathrm{Var}\,X \tag{4.5.4}$$

4.6 Equations

4.6.1 Cramer's Rule

Consider equation

$$\mathsf{A}\boldsymbol{x} = \boldsymbol{b} \tag{4.6.1}$$

where A is a $n \times n$ matrix, and $\boldsymbol{x}$ and $\boldsymbol{b}$ are $n \times 1$ vectors. Cramer's rule is the solution

$$x_i = \frac{|\mathsf{A}_i|}{|\mathsf{A}|} \tag{4.6.2}$$

where A_i is matrix A with i-th column replaced by $\boldsymbol{b}$.

4.7 Linear Algebra

4.7.1 Matrix

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

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

4.7.2 Determinant

Let

$$\mathsf{M} = \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{C} & \mathsf{D} \end{bmatrix} \tag{4.7.1}$$

Now

$$|\mathsf{M}| = |\mathsf{A} - \mathsf{B}\mathsf{D}^{-1}\mathsf{C}| \cdot |\mathsf{D}| \tag{4.7.2}$$

4.7.3 Inverse Matrix

Inverse matrix

$$\mathsf{A}^{-1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{|A|} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{4.7.3}$$

Partitioned inverse formula

$$\mathsf{M} = \begin{bmatrix} \mathsf{A} & \mathsf{B} \\ \mathsf{C} & \mathsf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathsf{E}^{-1} & -\mathsf{E}^{-1}\mathsf{B}\mathsf{D}^{-1} \\ -\mathsf{D}^{-1}\mathsf{C}\mathsf{E}^{-1} & \mathsf{F}^{-1} \end{bmatrix} \tag{4.7.4}$$

where

$$\begin{aligned} \mathsf{E} &= \mathsf{A} - \mathsf{B}\mathsf{D}^{-1}\mathsf{C} \equiv \mathsf{M}/\mathsf{D} \quad \text{is the Schur complement of } \mathsf{M} \text{ wrt } \mathsf{D} \\ \mathsf{E}^{-1} &= \mathsf{A}^{-1} + \mathsf{A}^{-1}\mathsf{B}\mathsf{F}^{-1}\mathsf{C}\mathsf{A}^{-1} \\ \mathsf{F} &= \mathsf{D} - \mathsf{C}\mathsf{A}^{-1}\mathsf{B} \equiv \mathsf{M}/\mathsf{A} \\ \mathsf{F}^{-1} &= \mathsf{D}^{-1} + \mathsf{D}^{-1}\mathsf{C}\mathsf{E}^{-1}\mathsf{B}\mathsf{D}^{-1} \end{aligned}$$

Proof: Murphy (2012, pp 118-119)

Moore-Penrose pseudoinverse

$$\mathsf{A}^{+} = \lim_{\alpha \downarrow 0} \left(\mathsf{A}^T\mathsf{A} + \alpha\mathsf{I}\right)^{-1} \mathsf{A}^T \tag{4.7.5}$$

M-P pseudoinvese solves the equation

$$y = \mathsf{A}x \quad \Rightarrow \quad x = \mathsf{A}^{+}y \tag{4.7.6}$$

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

4.7.4 Matrix norm

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

$$||\mathsf{A}||_F = \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{\mathrm{Tr}(\mathsf{A} \cdot \mathsf{A}^T)} \tag{4.7.7}$$

4.8 Matrix Calculus

4.8.1 Gradient of scalar function

Let $f : \mathbb{R}^K \to \mathbb{R}$ be a scalar-valued function on K-dimensional space $\mathbb{R}^K$. We denote this by $f(x)$ where $x \in \mathbb{R}^K$. Define the gradient

$$\nabla_x f(x) \equiv \frac{\partial f(x)}{\partial x} = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \dots \\ \frac{\partial}{\partial x_K} f(x). \end{bmatrix} \tag{4.8.1}$$

This is called *denominator layout* (Hessian formulation). Alternatively one can use *numerator layout* (Jacobian formulation):

$$\nabla_x f(x) \equiv \frac{\partial f(x)}{\partial x} = \left[\frac{\partial}{\partial x_1} f(x), \dots, \frac{\partial}{\partial x_K} f(x). \right] \tag{4.8.2}$$

Below we stay with denominator layout.

4.8.2 Chain Rule

Let $y = y(x)$ and $x = x(z)$ where y and z are scalars and x is a factors. The chain rule for the first derivative:

$$\frac{\partial y(x(z))}{\partial z} = \frac{\partial y(x)}{\partial x^{\mathsf{T}}} \cdot \frac{\partial x(z)}{\partial z} = \frac{\partial x(z)}{\partial z}^{\mathsf{T}} \cdot \frac{\partial y(x)}{\partial x}. \tag{4.8.3}$$

Note the first derivative is a scalar. See also the chain rule in scalar form in Section 3.4.6.

Chain rule for the second derivative:

$$\frac{\partial^2 y(x(z))}{\partial z^2} = \frac{\partial^2 x(z)}{z^2}^{\mathsf{T}} \cdot \frac{\partial y(x)}{\partial x} + \frac{\partial x(z)}{z}^{\mathsf{T}} \cdot \frac{\partial^2 y(x)}{\partial x^{\mathsf{T}} \partial x} \cdot \frac{\partial x(z)}{z}. \tag{4.8.4}$$

Note: this is a scalar.

4.8.3 Other Rules

Let $\boldsymbol{\beta}$ be a $K \times 1$ a vector. Then:

$$\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{x}^{\top}} = \frac{\partial \boldsymbol{x}^{\top}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial x_1}{\partial x_1} & \frac{\partial x_1}{\partial x_2} & \cdots & \frac{\partial x_1}{\partial x_K} \\ \frac{\partial x_2}{\partial x_1} & \frac{\partial x_2}{\partial x_2} & \cdots & \frac{\partial x_2}{\partial x_K} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_K}{\partial x_1} & \frac{\partial x_K}{\partial x_2} & \cdots & \frac{\partial x_K}{\partial x_K} \end{bmatrix} = \mathsf{I} \tag{4.8.5}$$

$$\frac{\partial \boldsymbol{x}^{\top}\boldsymbol{\beta}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{\beta}^{\top}\boldsymbol{x}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial}{\partial x_1}(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K) \\ \frac{\partial}{\partial x_2}(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K) \\ \vdots \\ \frac{\partial}{\partial x_K}(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K) \end{bmatrix} = \boldsymbol{\beta} \tag{4.8.6}$$

$$\frac{\partial\, \boldsymbol{\beta}^{\top}\boldsymbol{x}\boldsymbol{x}^{\top}\boldsymbol{\beta}}{\partial \boldsymbol{x}} = 2(\boldsymbol{x}^{\top}\boldsymbol{\beta}) \cdot \boldsymbol{\beta} \tag{4.8.7}$$

Let A be a $K \times N$ matrix and B $K \times K$ matrix

$$\frac{\partial \mathsf{A}\boldsymbol{x}}{\partial \boldsymbol{x}^{\top}} = \mathsf{A} \qquad \frac{\partial \boldsymbol{x}^{\top}\mathsf{A}^{\top}}{\partial \boldsymbol{x}} = \mathsf{A}^{\top} \tag{4.8.8}$$

$$\frac{\partial \boldsymbol{x}^{\top}\boldsymbol{x}}{\partial \boldsymbol{x}} = \mathsf{I}\boldsymbol{x} + \boldsymbol{x}\mathsf{I} = 2\boldsymbol{x} \qquad \frac{\partial \boldsymbol{x}^{\top}\mathsf{B}\boldsymbol{x}}{\partial \boldsymbol{x}} = (\mathsf{B} + \mathsf{B}^{\top})\boldsymbol{x} \tag{4.8.9}$$

Let $\boldsymbol{F} : \mathbb{R}^{N\times 1} \to \mathbb{R}^{M\times 1}$ and $\boldsymbol{x} \in \mathbb{R}^{N\times 1}$:

$$\frac{\partial}{\partial \boldsymbol{x}}[\boldsymbol{F}(\boldsymbol{x})^{\top} \cdot \boldsymbol{F}(\boldsymbol{x})] = 2\frac{\partial \boldsymbol{F}(\boldsymbol{x})^{\top}}{\partial \boldsymbol{x}} \cdot \boldsymbol{F}(\boldsymbol{x}) \tag{4.8.10}$$

Let

$$\mathsf{A} = \begin{pmatrix} a_{1\bullet}^{\top} \\ a_{2\bullet}^{\top} \\ \cdots \\ a_{k\bullet}^{\top} \\ \cdots \\ a_{N\bullet}^{\top} \end{pmatrix} \tag{4.8.11}$$

be a $N \times M$ matrix where $a_{i\bullet} = (a_{i1}, a_{i2}, \ldots, a_{ik}, \ldots, a_{iM})^{\top}$ is it's i-th row as $M \times 1$ column vector. Let B also be a $N \times M$ matrix. Let $\boldsymbol{z}$ be length-N vector $\boldsymbol{z} = (z_1, z_2, \ldots, z_N)^{\top}$. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be length-M vectors $\boldsymbol{x} = (x_1, x_2, \ldots, x_M)^{\top}$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_M)^{\top}$. Further, denote by $\boldsymbol{a}_{\bullet j}$ the j-th column of matrix A in for of $N \times 1$ column vector. Now

$$\frac{\partial}{\partial a_{j\bullet}^{\top}}\mathsf{A}\boldsymbol{a}_{k\bullet} = \delta_{kj}\mathsf{A} + \begin{pmatrix} \mathbf{0}^{\top} \\ \mathbf{0}^{\top} \\ \vdots \\ \boldsymbol{a}_{k\bullet}^{\top} \\ \vdots \\ \mathbf{0}^{\top} \end{pmatrix} \quad \text{and} \quad \frac{\partial}{\partial a_{j\bullet}^{\top}}\boldsymbol{z}^{\top}\mathsf{A}\boldsymbol{a}_{k\bullet} = \delta_{kj}\boldsymbol{z}^{\top}\mathsf{A} + z_j\boldsymbol{a}_{k\bullet}^{\top} \tag{4.8.12}$$

where $\mathbf{0}$ is a vector of zeros of suitable length and δ_{kj} is Kronecker delta.

$$\frac{\partial}{\partial \mathsf{A}} z^{\mathsf{T}} \mathsf{A} a_{k\bullet} = \begin{pmatrix} \mathbf{0}^{\mathsf{T}} \\ \mathbf{0}^{\mathsf{T}} \\ \vdots \\ z^{\mathsf{T}} \cdot \mathsf{A} \\ \vdots \\ \mathbf{0}^{\mathsf{T}} \end{pmatrix} + (\mathbf{1}_N \otimes a_{k\bullet}^{\mathsf{T}}) \odot (z \otimes \mathbf{1}_M^{\mathsf{T}}) \tag{4.8.13}$$

where $\otimes$ is Kronecker product, $\odot$ is element-wise product, and the the first matrix contains zeros everywhere except in the k-th row.

$$\frac{\partial}{\partial a_{ij}} x^{\mathsf{T}} \mathsf{A}^{\mathsf{T}} \mathsf{A} y = x_j (a_{i\bullet}^T \cdot y) + (x^T \cdot a_{i\bullet}) y_j \tag{4.8.14}$$

$$\frac{\partial}{\partial a_{i\bullet}} x^{\mathsf{T}} \mathsf{A}^{\mathsf{T}} \mathsf{A} y = x^{\mathsf{T}} \otimes (a_{i\bullet}^T \cdot y) + (x^T \cdot a_{i\bullet}) \otimes y \tag{4.8.15}$$

$$\frac{\partial}{\partial \mathsf{A}} x^{\mathsf{T}} \mathsf{A}^{\mathsf{T}} \mathsf{A} y = x \cdot y^{\mathsf{T}} \cdot \mathsf{A}^{\mathsf{T}} + y \cdot x^{\mathsf{T}} \cdot \mathsf{A}^{\mathsf{T}} \tag{4.8.16}$$

$$\frac{\partial}{\partial \mathsf{A}} x^{\mathsf{T}} \mathsf{A}^{\mathsf{T}} \mathsf{B} y = x \cdot y^{\mathsf{T}} \cdot \mathsf{B}^{\mathsf{T}} \tag{4.8.17}$$

$$\frac{\partial}{\partial \mathsf{B}} x^{\mathsf{T}} \mathsf{A}^{\mathsf{T}} \mathsf{B} y = y \cdot x^{\mathsf{T}} \cdot \mathsf{A}^{\mathsf{T}} \tag{4.8.18}$$

Note: as $a_{i\bullet}^T \cdot y$ and $x^T a_{i\bullet}$ are scalars, the Kronecker product in (4.8.15) are equivalent to scalar product.

$$\begin{aligned} \frac{\partial}{\partial \mathsf{A}} \left[\log \frac{1}{1 + \exp(-x^{\mathsf{T}} \mathsf{A}^{\mathsf{T}} \mathsf{A} y)} \right] &= \\ &= \frac{1}{1 + \exp(x^{\mathsf{T}} \mathsf{A}^{\mathsf{T}} \mathsf{A} y)} \cdot \left[x \cdot y^{\mathsf{T}} \cdot \mathsf{A}^{\mathsf{T}} + y \cdot x^{\mathsf{T}} \cdot \mathsf{A}^{\mathsf{T}} \right] \end{aligned} \tag{4.8.19}$$

Traces and such

$$\frac{\partial}{\partial \mathsf{A}} \operatorname{Tr}(\mathsf{BA}) = \mathsf{B}^{\mathsf{T}} \qquad \frac{\partial}{\partial \mathsf{A}} \log |\mathsf{A}| = \left(\mathsf{A}^{-1} \right)^{\mathsf{T}} \tag{4.8.20}$$

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

$$\frac{\partial}{\partial x} [f(x)' \mathsf{B} f(x)] = [D f(x)]' (\mathsf{B} + \mathsf{B}') f(x), \tag{4.8.21}$$

where $D f(x)$ 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:

$$\frac{\partial}{\partial \boldsymbol{\beta}'}\left[(\boldsymbol{\beta}'\boldsymbol{x}) \otimes \boldsymbol{y}\right] = \frac{\partial \boldsymbol{\beta}'\boldsymbol{x}}{\partial \boldsymbol{\beta}'} \otimes \boldsymbol{y} = \boldsymbol{x}' \otimes \boldsymbol{y} = \boldsymbol{y}\boldsymbol{x}'. \tag{4.8.22}$$

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

4.9 Võrratused

4.9.1 Hölder's inequality

Let X and Y be random variables.

$$\mathbb{E}\,|XY| \le \left\{\mathbb{E}\left[|X|^{\frac{1}{\alpha}}\right]\right\}^{\alpha} \left\{\mathbb{E}\left[|X|^{\frac{1}{1-\alpha}}\right]\right\}^{1-\alpha} \tag{4.9.1}$$

4.9.2 Jensen's Inequality

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

$$\sum_i \lambda_i f(\boldsymbol{x}_i) \le f\left(\sum_i \lambda_i \boldsymbol{x}_i\right), \qquad \sum \lambda_i = 1 \tag{4.9.2}$$

$$\mathbb{E}\, f(x) < f(\mathbb{E}\, x). \tag{4.9.3}$$

Proof: definition of concavity, induction.

4.9.3 Triangle Inequality

$$|x + y| \le |x| + |y| \tag{4.9.4}$$

4.9.4 Cauchy-Schwartzi võrratus

$$| < x, y > | \le ||x|| \cdot ||y|| \tag{4.9.5}$$

For sequences

$$\left(\sum a_i b_i\right)^2 \le \left(\sum a_i\right)^2 \left(\sum b_i\right)^2. \tag{4.9.6}$$

For functions:

$$< x, y >= \int x \cdot y \,\mathrm{d}x \qquad ||x|| = \sqrt{< x, x >}$$

Vektorkujul:

$$(\boldsymbol{a} \cdot \boldsymbol{b})^2 \le ||\boldsymbol{a}||^2 ||\boldsymbol{b}||^2 \tag{4.9.7}$$

ehk siis ka

$$\sum_i \boldsymbol{z}_i \boldsymbol{z}'_i \ge \frac{\sum_i a_i \boldsymbol{z}_i \sum_i a_i \boldsymbol{z}'_i}{\sum_i a_i^2} \tag{4.9.8}$$

4.9.5 Inequalities, containing exponent

Proof in most cases by analysing the corresponding function.

$$\mathrm{e}^a \ge a \tag{4.9.9}$$

$$1 - \mathrm{e}^{-a} < a \tag{4.9.10}$$

$$(1 - \mathrm{e}^{-a})\mathrm{e}^{-a} < a \tag{4.9.11}$$

$$(1 + a)\mathrm{e}^a \ge (1 + 2a) \tag{4.9.12}$$

$$(1 + a)\mathrm{e}^{-a} < 1 \qquad \text{if} \quad a > 0 \tag{4.9.13}$$

$$(a - 1)\mathrm{e}^a > -1 \qquad \text{if} \quad a > 0 \tag{4.9.14}$$

$$(1 + a^2)\mathrm{e}^{-a} < 1 \qquad \text{if} \quad a > 0 \tag{4.9.15}$$

$$\mathrm{e}^{-a} - \mathrm{e}^{-b} < -a + b \qquad \text{if} \quad 0 < a < b \tag{4.9.16}$$

$$\mathrm{e}^{a} - \mathrm{e}^{b} = (a - b) + \frac{1}{2}(a^2 - b^2) + \frac{1}{6}(a^3 - b^3) + \dots \tag{4.9.17}$$

$$\gtrless (a - b) \qquad \text{if} \quad a \gtrless b \tag{4.9.18}$$

$$a\mathrm{e}^{-b} - b\mathrm{e}^{-a} \gtrless a - b \qquad \text{if} \quad a \gtrless b \tag{4.9.19}$$

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

$$\text{if} \quad b \gtrless a \tag{4.9.20}$$

4.10 Differential Equations

4.10.1 Linear Equations with Constant Coefficients

Homogeneous linear differential equations with constant coefficients are of form

$$a_n y^{(n)} + a_{n-1} y^{(n-1)} + \dots + a_1 y = 0 \tag{4.10.1}$$

It's particular solution is in the form

$$y = \mathrm{e}^{cx} \tag{4.10.2}$$

where c is a root of the characteristic polynomial

$$a_n c^n + a_{n-1} c^{n-1} + \dots + a_1 c = 0. \tag{4.10.3}$$

The general solution is

$$y(x) = u_1 \mathrm{e}^{c_1 x} + u_2 \mathrm{e}^{c_2 x} + \dots + u_n \mathrm{e}^{c_n x} \tag{4.10.4}$$

where $c_i, i = 1 \dots n$, are the characteristics roots of (4.10.3) and u_i are constants to be determined.

Proof: insert $y(x) = \mathrm{e}^{cx}$ into (4.10.1) and take derivatives.

First order linear differential equation has a general form

$$y'(x) + p(x)y(x) = q(x) \tag{4.10.5}$$

It's solution is

$$y(x) = \frac{\int u(x)q(x)\mathrm{d}x + C}{u(x)} \tag{4.10.6}$$

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

$$u(x) = \mathrm{e}^{\int p(x)\mathrm{d}x} \tag{4.10.7}$$

5 Probability and Statistics

5.1 Combinatorics

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

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

5.2 Basic statistics

Mean Let $\boldsymbol{x}$ be a $N \times 1$ vector. Its mean

$$\bar{x} = \frac{1}{N}\mathbf{1}^\top \boldsymbol{y} \tag{5.2.1}$$

Its deviations from the mean

$$\boldsymbol{x} - \bar{x} = \mathsf{D}\,\boldsymbol{x} = \left(\mathsf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top\right)\boldsymbol{x} \tag{5.2.2}$$

This is a conclusion from the linear regression residuals' formula (9.1.5). Its variance

$$\operatorname{Var}\boldsymbol{x} = \frac{1}{N}(\boldsymbol{x} - \bar{x})^\top(\boldsymbol{x} - \bar{x}) = \frac{1}{N}\boldsymbol{x}^\top\mathsf{D}^\top\mathsf{D}\,\boldsymbol{x} = \frac{1}{N}\boldsymbol{x}^\top\mathsf{D}^2\,\boldsymbol{x} \tag{5.2.3}$$

as D is symmetric.

5.3 Random Variables

5.3.1 General Concepts

Variance

$$\operatorname{Var}X = \mathbb{E}(X - \mathbb{E}\,X)^2 \tag{5.3.1}$$

Other Moments properties:

$$\mathbb{E}(X - \mathbb{E}\,X + \alpha)^2 = \operatorname{Var}X + \alpha^2 \tag{5.3.2}$$

$$\mathbb{E}(X - \mathbb{E}\,X + \alpha)^3 = \mathbb{E}(X - \mathbb{E}\,X)^3 + 3\alpha\operatorname{Var}X + \alpha^3 \tag{5.3.3}$$

where α 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 $F_X(x) \le F_Y(x) \quad \forall x$. Intuitively this means Y realizations are larger than X realizations.

5.3.2 Information and Entropy

Entropy of random variable X describes how much information we will gain from an observation of X, and is defined as

$$\mathbb{H}(X) = -\,\mathbb{E}\log f_X \tag{5.3.4}$$

For discrete distribution

$$\mathbb{H}(X) = -\sum_k \Pr(X = k) \log \Pr(X = k). \tag{5.3.5}$$

In case of discrete uniform with K possible states, $\mathbb{H}(X) = \log K$. For continuous distribution

$$\mathbb{H}(X) = -\int f_X(x) \cdot \log f_X(x)\, \mathrm{d}x. \tag{5.3.6}$$

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

Cross entropy

$$\mathbb{H}(p, q) = -\sum_k p_k \log q_k \tag{5.3.7}$$

5.3.3 Kullback-Leibler divergence

A measure of dissimilarity between distributions p and q:

$$KL(p||q) = \mathbb{E}_p \left[\log \frac{p(X)}{q(X)}\right] = \sum_k p_k \log \frac{p_k}{q_k} \tag{5.3.8}$$

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 a metric distance as $KL(p||q) \neq KL(q||p)$.

5.3.4 Mutual Information

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

$$MI(X, Y) = KL(P_{XY}(X, Y)||P_X(X)P_Y(Y)) = \sum_x \sum_y p(x, y) \log \frac{P_{XY}(X, Y)}{P_X(X)P_Y(Y)}. \tag{5.3.9}$$

Properties:

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

5.3.5 Sufficient statistics

$s(\boldsymbol{X})$ is a *sufficient statistics* of $\boldsymbol{X}$ if

$$P(\boldsymbol{\theta}|\boldsymbol{X}) = P(\boldsymbol{\theta}|s(\boldsymbol{X})) \tag{5.3.10}$$

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

5.4 Stochastic Boundedness and Convergence in Probability

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

$$\Pr(|X_n/a_n| > M) < \epsilon \quad \forall n > N \qquad \text{or alternatively:} \qquad X_n = O_p(a_n) \tag{5.4.1}$$

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

$$\lim_{n\to\infty} \Pr(|X_n| \geq \epsilon) = 0 \quad \forall\epsilon > 0 \qquad \text{or alternatively:} \qquad X_n = o_p(1) \tag{5.4.2}$$

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

$$\forall\theta \in \Theta, \epsilon > 0, \eta > 0 \quad \exists N : \Pr(|Q_n(\theta) - Q(\theta)| \geq \epsilon) < \eta \quad \forall n > N \tag{5.4.3}$$

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

$$\sup_{\theta\in\Theta} |Q_n(\theta) - Q(\theta)| = o_p(1) \tag{5.4.4}$$

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

$$\Pr(|\Delta_n(\epsilon, \eta)| > \epsilon) < \eta \quad \forall n > n_0(\epsilon, \eta), \tag{5.4.5}$$

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

$$\sup_{\tilde{\theta}\in\mathcal{N}(\theta,\epsilon,\eta)} \left|\hat{Q}_n(\tilde{\theta}) - \hat{Q}_n(\theta)\right| < \Delta_n(\epsilon, \eta) \quad \forall n > n_0(\epsilon, \eta) \tag{5.4.6}$$

See Newey (1991)

5.5 Distributions: General Concepts

5.5.1 Sõltumatud juhuslikud muutujad

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

5.5.2 Expectations

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

$$\mathbb{E}\,X = \int_a^b x\,\mathrm{d}F_X(x) \tag{5.5.1}$$

$$= a + \int_a^b \bar{F}_X(x)\,\mathrm{d}x. \tag{5.5.2}$$

Law of iterated expectations

$$\mathop{\mathbb{E}}_{X}\left[\mathbb{E}[Y|X]\right] = \mathbb{E}[Y] \tag{5.5.3}$$

5.5.3 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:

$$\mu_1 \equiv \mu \tag{5.5.4}$$

$$\sigma^2 \equiv m_2 = \mu_2 - \mu^2 \tag{5.5.5}$$

$$m_3 = \mu_3 - 3\mu\mu_2 + 2\mu^3 \tag{5.5.6}$$

$$m_4 = \mu_4 - 4\mu_3\mu + 6\mu_2\mu^2 - 3\mu^4 \tag{5.5.7}$$

5.5.4 Characteristic Function

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

$$\phi_X(t) = \mathbb{E}\, \mathrm{e}^{itX} \tag{5.5.8}$$

Properties: for independent random variables X_1, X_2

$$\phi_{X_1+X_2}(t) = \phi_{X_1}(t) \cdot \phi_{X_2}(t) \tag{5.5.9}$$

5.5.5 Momendifunktsioon (*moment generating function*)

MGF avaldub ühemõõtmelisel juhul:

$$M_x(s) = \mathbb{E}\, e^{sx} = \int e^{sx} f(x) dx. \tag{5.5.10}$$

Momendifunktsiooni omadused:

$$M'_x(0) = \mathbb{E}\, x \qquad M''_x(0) = \mathbb{E}\, x^2 \qquad M^{(n)}_x(0) = \mathbb{E}\, x^n. \tag{5.5.11}$$

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

$$M_{\log x}(s) = \mathbb{E}\, \mathrm{e}^{s \log x} = \mathbb{E}\, x^s. \tag{5.5.12}$$

N-mõõtmelisel juhul avaldub MGF:

$$\begin{aligned} M(s_1, s_2, \ldots, s_N) &= \mathbb{E}\, e^{\sum^N s_i x_i} = \\ &= \int \cdots \int e^{s_1 x_1} e^{s_2 x_2} \ldots e^{s_N x_N} f(x_1, x_2, \ldots, x_N) dx_1 dx_2 \ldots dx_N. \end{aligned} \tag{5.5.13}$$

5.5.6 Kumulandifunktsoon (*cumulant-generating function*)

KGF avaldub MGF-i kaudu:

$$K_x(s) = \log M(s) \qquad \text{või} \qquad K(s_1, s_2, \ldots, s_N) = \log M(s_1, s_2, \ldots, s_N). \tag{5.5.14}$$

KGF-i omadus (ühemõõtmelisel juhul):

$$K_x'(0) = \mathbb{E}\,x \tag{5.5.15}$$
$$K_x''(0) = \operatorname{Var} x \tag{5.5.16}$$
$$K_x'''(0) = \mathbb{E}(x - \mathbb{E}\,x)^3 \tag{5.5.17}$$

ja kahemõõtmelisel juhul:

$$\frac{\partial^2 K(0,0)}{\partial s_1 \partial s_2} = \operatorname{Cov}(x_1, x_2). \tag{5.5.18}$$

5.5.7 Probability Generating Function

For discrete, non-negative random variables

$$G(z) = \mathbb{E}(z^X) = \sum_{x=0}^{\infty} p(x) z^x. \tag{5.5.19}$$

Properties:

$$p(k) = \Pr(X = k) = \frac{G^{(k)}(0)}{k!} \tag{5.5.20}$$

5.5.8 Distribution Function of a Function of Random Variable

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

$$F_x(x) = \Pr[X < x] = \Pr[X < x(y)] = F_x[x(y)] = F_y(y) \tag{5.5.21}$$

ja

$$f_y(y) = F_y'(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_x[x(y)] = \frac{\mathrm{d}}{\mathrm{d}x} F_x(x) \frac{\mathrm{d}x}{\mathrm{d}y} = f_x[x(y)] \frac{\mathrm{d}x}{\mathrm{d}y}. \tag{5.5.22}$$

5.6 One-dimensional discrete distributions

5.6.1 Bernoulli

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

$$X = \mathbb{1}(E) = \begin{cases} 1 & \text{if } E, \\ 0 & \text{if } \bar{E}. \end{cases} \tag{5.6.1}$$

has Bernoulli distribution.

Properties:

$$\mathbb{E}\, X = \mu \tag{5.6.2}$$
$$\operatorname{Var} X = \mu(1-\mu) \tag{5.6.3}$$

Exponential family minimal canonical parameters

$$\eta = \log\left(\frac{\mu}{1-\mu}\right) \qquad \mu = \Lambda(\eta) \tag{5.6.4}$$
$$A(\eta) = \log(1+\mathrm{e}^{\eta}) = \log\left(\frac{1}{1-\mu}\right) \tag{5.6.5}$$
$$h(x) = 1 \tag{5.6.6}$$
$$T(x) = x \tag{5.6.7}$$

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

5.6.2 Binoomjaotus

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

$$\begin{aligned}
\Pr(X=x) &= \binom{N}{x} p^x (1-p)^{N-x} \qquad x \in \{0,\dots,N\} \\
\mathbb{E}\, X &= Np \\
\mathbb{E}(X-\mathbb{E}\, X)^2 &= Np(1-p) \\
\mathbb{E}(X-\mathbb{E}\, X)^3 &= Np(1-p)(1-2p) \\
\mathbb{E}(X-\mathbb{E}\, X)^4 &= Np(1-p)(1-3p+3p^2) \\
\mathbb{E}(X-\mathbb{E}\, X)^5 &= Np(1-p)(1-2p)(1-2p+2p^2) \\
\mathbb{E}(X-\mathbb{E}\, X)^6 &= Np(1-p)(1-5p+10p^2-10p^3+5p^4)
\end{aligned}$$

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

5.6.3 Diskreetne jaotus

Jaotus kus juhslikul muutujal võib olla lõplik hulk diskreetseid väärtusi.

5.6.4 Generalized Poisson

It is a generalization of Poisson distribution, it allows overdispersion.

Properties

$$\text{pdf} \quad f(n;\lambda,\eta) = \frac{\lambda(\lambda+\eta n)^{n-1}\,\mathrm{e}^{-\lambda-\eta n}}{n!} \tag{5.6.8}$$
$$\mathbb{E}\, n = \frac{\lambda}{1-\eta} \tag{5.6.9}$$

$$\mathrm{Var}\, n = \frac{\lambda}{(1-\eta)^3} \tag{5.6.10}$$

$$\text{overdispersion} \quad D = \frac{\mathbb{E}\, n}{\sqrt{\mathrm{Var}\, n}} = \frac{1}{\sqrt{\lambda(1-\eta)}} \tag{5.6.11}$$

- GP is a mixture of Poisson distribution (Joe and Zhu, 2005).

Tuenter (2000) shows that $\sum_{n=0}^{\infty} = 1$.

Alternative parameterizations

Scale and shape Define scale (mean) $\mu = \frac{\lambda}{1-\eta} > 0$ and shape $\alpha = \frac{\eta}{\lambda}$. α measures the deviation from standard Poisson (where $\alpha = 0$), measured in expected value. (The reverse transformation $\lambda = \frac{\mu}{1+\alpha\mu}$ and $\eta = \frac{\alpha\mu}{1+\alpha\mu}$.) Now

$$\text{pdf} \quad f(n;\mu,\alpha) = \left(\frac{n}{1+\alpha n}\right)^n \frac{(1+\alpha n)^{n-1}}{n!} \exp\left(-n\frac{\mu(1+\alpha n)}{1+\alpha}\right) \tag{5.6.12}$$

$$\mathbb{E}\, n = \mu \tag{5.6.13}$$

$$\mathrm{Var}\, n = \mu(1+\alpha\mu)^2 \tag{5.6.14}$$

$$\text{overdispersion} \quad D = \frac{1+\alpha\mu}{\sqrt{\mu}} \tag{5.6.15}$$

Mean and variance Let μ be the expectation and σ^2 the variance. The reverse transformation $\lambda = \frac{\mu^{3/2}}{\sigma}$ and $\eta = 1 - \frac{\mu^{1/2}}{\sigma}$.

5.6.5 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:

$$f(n) = (1-p)p^n. \tag{5.6.16}$$

Jaotuse omadused:

$$\mathbb{E}\, n = \frac{1-p}{p} \tag{5.6.17}$$

$$\mathrm{Var}\, n = \frac{1-p}{p^2} \tag{5.6.18}$$

5.6.6 Multinoomjaotus

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

$$\mathbb{E}\, N_i = Np_i \tag{5.6.19}$$

$$\operatorname{Var} N_i = Np_i(1 - p_i) \tag{5.6.20}$$

$$\dots$$

$$\operatorname{Cov}(N_i, N_j) = -Np_ip_j \tag{5.6.21}$$

Tõestus: momentide arvutamisel võib multinoomjaotuse taandada binoomjaotuseks, kovariatsiooni jaoks kirjuta definitsioon lahti, arvesta et $EA_iA_j = 0$.

5.6.7 Negative Binomial

Negative binomial is an overdispersed distribution counts. It arises as a number of failures in Bernoulli process for a given number of successes. Let n be the number of failures F till we get to s successes S. Let p be the probability of success. The probability to acheve s-th success after n failures is $\Pr(S = s-1, F = n)\Pr(S) = \binom{s+n-1}{s-1} p^{s-1}(1-p)^n p = \binom{s+n-1}{s-1} p^s(1-p)^n$.

Properties:

$$\text{pmf} \quad f(n; s, p) = \binom{s+n-1}{s-1} p^s(1-p)^n \tag{5.6.22}$$

$$= \frac{\Gamma(s+n)\, p^s(1-p)^n}{\Gamma(s)\, n!} \quad \text{for continuous } s \tag{5.6.23}$$

$$\mathbb{E}\, n = \frac{s(1-p)}{p} \tag{5.6.24}$$

$$\operatorname{Var} n = \frac{s(1-p)}{p^2} \tag{5.6.25}$$

$$\text{mode} = \left\lfloor \frac{s(1-p)-1}{p} \right\rfloor + 1 \tag{5.6.26}$$

1. Negative binomial is a Poisson mixture where the mixing distribution $\lambda \sim Gamma(\theta, p/(1-p))$.

5.6.8 Poisson

Poisson distribution describes a sum of independent rare events, its pdf describes the probability to observe n events with its parameter, λ, describing the expected number of events. This can be applied to different time periods, if the expectations is μ events per time unit, then for time period t, the expected number (Poisson parameter) is $\lambda = \mu t$.

Properties:

$$\text{pdf} \quad f(n) = \frac{\lambda^n\, e^{-\lambda}}{n!} \tag{5.6.27}$$

$$\mathbb{E}\, n = \lambda \tag{5.6.28}$$

$$\operatorname{Var} n = \lambda \tag{5.6.29}$$

- Poisson pdf is log-concave
- Sum of Poisson variables is a Poisson variable:

$$\text{if} \quad X \sim Pois(\lambda_1), Y \sim Pois(\lambda_2) \quad \text{then} \quad X + Y \sim Pois(\lambda_1 + \lambda_2) \tag{5.6.30}$$

- Difference of two Poisson variables is Skellam distributed:

$$\text{if} \quad X \sim Pois(\lambda_1), Y \sim Pois(\lambda_2) \quad \text{then} \quad X - Y \sim Skellam(\lambda_1, \lambda_2) \tag{5.6.31}$$

Characteristic function:

$$\phi(t) = \mathrm{e}^{\lambda(\mathrm{e}^{it} - 1)} \tag{5.6.32}$$

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

$$\hat{\lambda} = \frac{\sum n_i}{\sum t_i} \qquad \text{ja} \qquad \operatorname{Var} \hat{\lambda} = \frac{\sum n_i}{(\sum t_i)^2} \tag{5.6.33}$$

Poissoni summa tuletis aja järgi: Olgu

$$\vartheta(t) = \sum_{s=0}^{S} Q(s) \frac{(\lambda t)^2}{s!} \mathrm{e}^{-\lambda t} = \mathop{\mathbb{E}}_{s} Q(s) \tag{5.6.34}$$

siis

$$\frac{\partial}{\partial t} \vartheta(t) = \lambda \sum_{s=0}^{S-1} \left[Q(s+1) - Q(s) \right] p_p(s) - \lambda Q(S) p_p(S) = \tag{5.6.35}$$

$$= \lambda \sum_{s=1}^{S} Q(s) \left[p_p(s-1) - p(s) \right] - \lambda Q(0) p_p(0) \tag{5.6.36}$$

$$\frac{\partial}{\partial t} P_p(s) = -\lambda p_p(s) \tag{5.6.37}$$

$$\frac{\partial}{\partial t} p_p(s) = \begin{cases} -\lambda p_p(s) & \text{kui } s = 0 \\ -\lambda p_p(s) + \lambda p_p(s-1) & \text{kui } s > 0 \end{cases} \tag{5.6.38}$$

Poisson distribution can be generalized to generalized Poisson distribution.

5.6.9 Skellam Distribution $PD(\lambda, \delta)$

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

$$P(N = n) = \mathrm{e}^{-(\lambda+\delta)} \left(\frac{\lambda}{\delta} \right)^{\frac{n}{2}} I_n(2\sqrt{\lambda\delta}) \tag{5.6.39}$$

$$\mathbb{E}\, N = \lambda - \delta \tag{5.6.40}$$

$$\operatorname{Var} N = \lambda + \delta \tag{5.6.41}$$

5.7 1D Continuous Distributions

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

$$\text{pdf} \qquad f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} \quad x \in [0,1] \tag{5.7.1}$$

$$\mathbb{E}\, X = \frac{a}{(a+b)} \tag{5.7.2}$$

$$\operatorname{Var} X = \frac{ab}{(a+b)^2(a+b+1)} \tag{5.7.3}$$

$$\text{mode} \qquad m = \frac{a-1}{a+b-2} \tag{5.7.4}$$

where $B(a, b)$ is is the beta function (2.3.11).

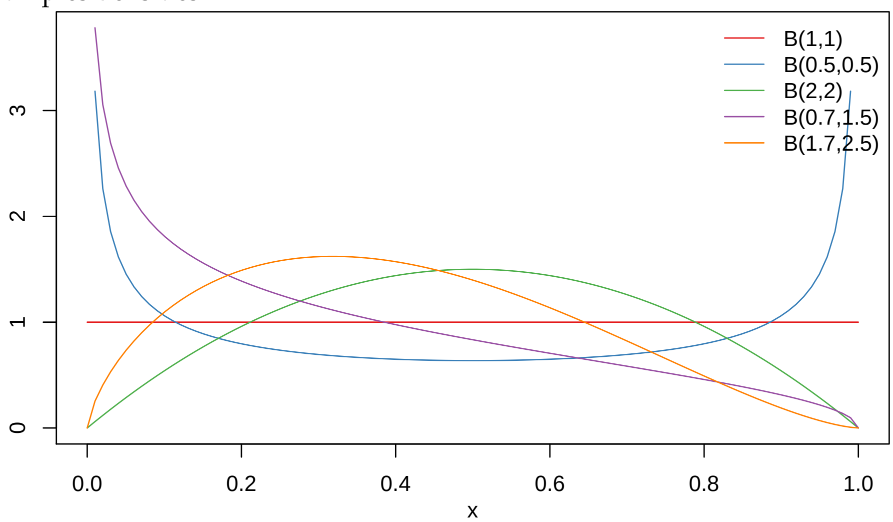
Moments:

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. Larger parameter values concentrate the probability mass more toward the center, smaller ones to the extremes. If parameters are unequal, the mass tends more toward the smaller value.

Multivariate generalization of beta distribution is Dirichlet distribution.

Examples densities:



5.7.2 Cauchy distribution

Cauchy distribution is a fat-tailed bell curve-like distribution. It is a special case of t-distribution with degrees of freedom = 1.

$$\text{pdf} \qquad f(x) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x-x_0}{\gamma}\right)^2\right]} \tag{5.7.5}$$

$$\text{mode} \qquad m = x_0 \tag{5.7.6}$$

$$\mathbb{E}\,X \quad \text{undefined} \tag{5.7.7}$$

where x_0 is location and γ is scale.

5.7.3 Eksponentjaotus $\mathcal{E}(\theta)$

Eksponentjaotus kirjeldab konstantse kiirusega hääbuvaid protsesse. Tõenäosustihedus:

$$f(t) = \theta e^{-\theta t}, \qquad t \geq 0, \theta > 0 \tag{5.7.8}$$

jaotusfunktsioon:

$$F(t) = 1 - e^{-\theta t}, \tag{5.7.9}$$

momendifunktsioon:

$$M_T(s) = \frac{1}{1 - \frac{s}{\theta}}, \qquad s < \theta. \tag{5.7.10}$$

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:

$$M_{\log T}(s) = \frac{\Gamma(s+1)}{\theta^s} \tag{5.7.11}$$

$$K_{\log T}(s) = \log\Gamma(s+1) - s\log\theta \tag{5.7.12}$$

$$\mathbb{E}\log T = \psi(1) - \log\theta \tag{5.7.13}$$

$$\text{Var}\,\log T = \psi'(1), \tag{5.7.14}$$

kus ψ on digamma funktsioon.

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

$$\log z_1 - \log z_2 \sim \frac{\theta_1}{\theta_1 + \theta_2 \mathrm{e}^{-x}} \sim \Lambda(x), \qquad \text{kui} \quad \theta_1 = \theta_2 \tag{5.7.15}$$

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

5.7.4 F-jaotus $F(n_1, n_2)$

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

$$\frac{\frac{w_1}{n_1}}{\frac{w_2}{n_2}} \sim F(n_1, n_2). \tag{5.7.16}$$

Tihedusfunktsioon:

$$f(x) = \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} x^{\frac{n_1}{2}-1}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(1 + \frac{n_1}{n_2}x\right)^{\frac{n_1+n_2}{2}}} \tag{5.7.17}$$

5.7.5 Gamma distribution $\mathcal{G}(\alpha, \beta)$

Is sometimes used to describe the unobserved heterogeneity in duration models.

$$\text{pdf} \quad f(x) = \frac{1}{\beta^\alpha} \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \qquad x > 0, \tag{5.7.18}$$

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

Properties

$$\begin{aligned}
\mathbb{E}\, X &= \beta\alpha && (5.7.19)\\
\mathbb{E}\, X^2 &= \beta^2\alpha(\alpha+1) && (5.7.20)\\
\operatorname{Var} X &= \beta^2\alpha && (5.7.21)\\
M_X(s) &= \frac{1}{(\beta s - 1)^\alpha} && (5.7.22)\\
K_X(s) &= -\alpha \log(\beta s - 1) && (5.7.23)\\
& && (5.7.24)
\end{aligned}$$

Properties related to $\log x$

$$\begin{aligned}
\mathbb{E} \log x &= \log \beta + \psi(\alpha) && (5.7.25)\\
\operatorname{Var} \log x &= \psi'(\alpha) && (5.7.26)\\
M_{\log x}(s) &= \frac{\Gamma(s+\alpha)}{\Gamma(\alpha)} \beta^s && (5.7.27)\\
K_{\log x}(s) &= s \log \beta + \log \Gamma(s+\alpha) - \log \Gamma(\alpha) && (5.7.28)\\
& && (5.7.29)
\end{aligned}$$

Special cases *Exponential distribution*: Gamma$(1, 1/\beta)$ is the same as Exp(β).

normaalne gammajaotus, mille keskväärtus on 1. Sel juhul

$$\alpha = \frac{1}{\beta} \equiv \eta \tag{5.7.30}$$

ja jaotusfunktsioon

$$f_x(x) = \eta^\eta \frac{1}{\Gamma(\eta)} x^{\eta-1} \mathrm{e}^{-\eta x}. \tag{5.7.31}$$

Sel juhul:

$$\begin{aligned} \mathbb{E}\, x &= 1 && (5.7.32)\\ \mathbb{E}\, x^2 &= 1 + \frac{1}{\eta} && (5.7.33)\\ \operatorname{Var} x &= \frac{1}{\eta} && (5.7.34) \end{aligned}$$

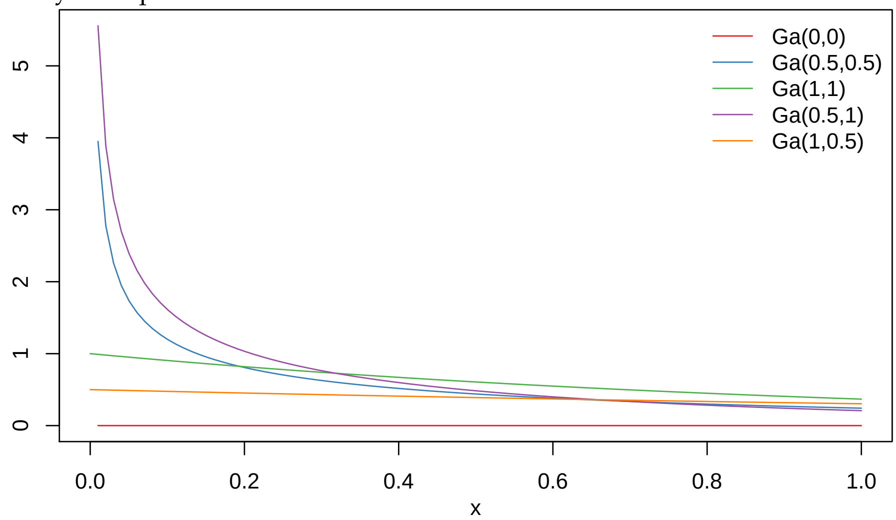
Teine oluline erijuht on χ^2-jaotus. Kui $\alpha = \frac{k}{2}$ ja $\beta = 2$, siis X jaotusfunktsioon on

$$f_x(x) = \frac{1}{2^{\frac{k}{2}}} \frac{1}{\Gamma(\frac{k}{2})} y^{\frac{k}{2}-1} \mathrm{e}^{-\frac{y}{2}}. \tag{5.7.35}$$

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

Gamma distribution generalized to matrices is Wishart distribution.

Density examples:



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

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

$$\begin{aligned} &\text{pdf} && f_x(x) = \frac{1}{2^{\frac{k}{2}}} \frac{1}{\Gamma(\frac{k}{2})} y^{\frac{k}{2}-1} \mathrm{e}^{-\frac{y}{2}} && (5.7.36)\\ &\text{MGF} && M(s) = (1-2s)^{-\frac{k}{2}} && (5.7.37) \end{aligned}$$

5.7.7 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

$$\text{density} \quad f(x) = \frac{\beta^{-\alpha}}{\Gamma(\alpha)} x^{-(\alpha+1)} \mathrm{e}^{-\frac{1}{\beta x}} \tag{5.7.38}$$

5.7.8 Laplace distribution

Also double-sided exponential distribution. Density:

$$f_X(x;\mu,b) = \frac{1}{2b}\mathrm{e}^{-\frac{|x-\mu|}{b}}. \tag{5.7.39}$$

Properties:

$$\mathbb{E}\,X = \mu \qquad \operatorname{Var} X = 2b^2 \tag{5.7.40}$$

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

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

$$\text{cdf} \qquad f_x(x) = \frac{1}{\sqrt{2\pi}\sigma x}\mathrm{e}^{-\frac{1}{2}\left[\frac{(\log x-\mu)}{\sigma}\right]^2} \tag{5.7.41}$$

Properties:

$$\mathbb{E}\,x = \mathrm{e}^{\mu+\frac{1}{2}\sigma^2} \tag{5.7.42}$$

$$\operatorname{Var} x = \mathrm{e}^{2\mu+\sigma^2}(\mathrm{e}^{\sigma^2}-1) \tag{5.7.43}$$

$$\text{median} = \mathrm{e}^{\mu} \tag{5.7.44}$$

5.7.10 Log-ühtlane jaotus

Kasutatakse palgajaotuse kirjeldamiseks.

$$\text{pdf} \qquad f(x) = \frac{1}{x}\cdot\frac{1}{\log\beta - \log\alpha} \quad \text{where} \quad 0 \le \alpha \le x \le \beta < \infty. \tag{5.7.45}$$

5.7.11 Logistic Distribution

$$\text{cdf} \qquad \Lambda(x) = \frac{e^x}{1+e^x} = \frac{1}{1+\mathrm{e}^{-x}} \tag{5.7.46}$$

$$\text{pdf} \qquad f(x) = \frac{\mathrm{e}^x}{(1+\mathrm{e}^x)^2} = \frac{\mathrm{e}^{-x}}{(1+\mathrm{e}^{-x})^2} =$$

$$= \Lambda(-x)\Lambda(x) = [1-\Lambda(x)]\Lambda(x) \tag{5.7.47}$$

$$f'(x) = \mathrm{e}^{-x}\frac{\mathrm{e}^{-x}-1}{(\mathrm{e}^{-x}+1)^3} \tag{5.7.48}$$

$$\text{MGF} \qquad M(s) = \int \frac{\mathrm{e}^{x(s+1)}}{(1+\mathrm{e}^x)^2}\mathrm{d}x \tag{5.7.49}$$

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

5.7.12 Lomax Distribution

(also Pareto II distribution)

Describes the upper part of many highly unequal distributions where the values can start from 0. It is a shift of Pareto distribution. Let x_0 be scale. Properties:

$$f_X(x) = \frac{\alpha}{x_0}\left(\frac{x}{x_0} + 1\right)^{-\alpha-1} \tag{5.7.50}$$

$$\mathbb{E}\,X = \frac{1}{\alpha - 1}x_0, \quad \alpha > 1 \tag{5.7.51}$$

Properties:

- power law: $\log f_X(x)$ 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.

Shifted Lomax, where the x values start from x_0, is Pareto Distribution .

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

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

$$F(x;\mu,\sigma) \equiv \Phi\left(\frac{x-\mu}{\sigma}\right) \qquad \text{cannot be expressed analytically} \tag{5.7.52}$$

$$f(x;\mu,\sigma) \equiv \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma}\mathrm{e}^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \tag{5.7.53}$$

Moments (from wikipedia):

Order	Non-central moment	Central moment
1	μ	0
2	$\mu^2 + \sigma^2$	σ^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:

$$\phi_X(t) = \mathrm{e}^{it\mu - \frac{1}{2}\sigma^2 t^2} \tag{5.7.54}$$

Moment generating function

$$M_x(s) = \mathrm{e}^{\mu s + \frac{1}{2}\sigma^2 s^2} \tag{5.7.55}$$

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

$$\sum_i X_i \sim N\left(\sum_i \mu_i, \sum_i \sigma_i^2\right). \tag{5.7.56}$$

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

$$\mathbb{E}[X|X < a] = \mu - \sigma \frac{\phi(\frac{a-\mu}{\sigma})}{\Phi(\frac{a-\mu}{\sigma})} = \mu - \sigma\lambda\left(\frac{a-\mu}{\sigma}\right) \tag{5.7.57}$$

$$\mathbb{E}[X|X > a] = \mu + \sigma \frac{\phi(\frac{a-\mu}{\sigma})}{1-\Phi(\frac{a-\mu}{\sigma})} = \mu + \sigma\lambda\left(\frac{\mu-a}{\sigma}\right) \tag{5.7.58}$$

$$\mathbb{E}[X|X \in [a,b]] = \mu - \sigma \frac{\phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} \tag{5.7.59}$$

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

$$\mathbb{E}[X|X < a] = -\sigma\lambda(\frac{a}{\sigma}) \tag{5.7.60}$$

$$\mathbb{E}[X|X > a] = \sigma\lambda(-\frac{a}{\sigma}) \tag{5.7.61}$$

$$\mathbb{E}[X^2|X < a] = \sigma^2 - \sigma a\lambda\left(\frac{a}{\sigma}\right) \tag{5.7.62}$$

$$\mathbb{E}[X^2|X > a] = \sigma^2 + \sigma a\lambda\left(-\frac{a}{\sigma}\right) \tag{5.7.63}$$

$$\mathbb{E}[X^2|X > -a \wedge X < a] = \sigma^2 - 2\frac{\sigma a\phi\left(\frac{a}{\sigma}\right)}{1-2\Phi\left(\frac{a}{\sigma}\right)} \tag{5.7.64}$$

$$\mathbb{E}[X^2|X < -a \vee X > a] = \mathbb{E}[X^2|X > a] \tag{5.7.65}$$

$$\mathrm{Var}\,[X|X < a] = \sigma^2\left[1 - \frac{a}{\sigma}\lambda\left(\frac{a}{\sigma}\right) - \lambda^2(\frac{a}{\sigma})\right] \tag{5.7.66}$$

$$\mathrm{Var}\,[X|X > a] = \sigma^2\left[1 + \frac{a}{\sigma}\lambda\left(-\frac{a}{\sigma}\right) - \lambda^2(-\frac{a}{\sigma})\right] \tag{5.7.67}$$

$$\tag{5.7.68}$$

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

$$\mathbb{E}[X|X < Y] = \mu_X - \frac{\sigma_X^2}{\sqrt{\sigma_X^2 + \sigma_Y^2}}\lambda\left(-\frac{\mu_X - \mu_Y}{\sqrt{\sigma_X^2 + \sigma_Y^2}}\right) \tag{5.7.69}$$

$$\mathbb{E}[X|X > Y] = \mu_X + \frac{\sigma_X^2}{\sqrt{\sigma_X^2 + \sigma_Y^2}}\lambda\left(\frac{\mu_X - \mu_Y}{\sqrt{\sigma_X^2 + \sigma_Y^2}}\right). \tag{5.7.70}$$

Proof: write $\mathbb{E}[X|X < Y] = \mathbb{E}[X|Z < 0]$ where $Z = X - Y$. Now follows from (5.8.13)

5.7.14 Pareto Distribution

Describes the upper part of many highly unequal distributions.

$$F_X(x) = 1 - \left(\frac{x_0}{x}\right)^\alpha, \quad x \geq x_0 > 0; \quad \alpha > 0 \tag{5.7.71}$$

$$f_X(x) = \alpha\, x_0^{\alpha}\, x^{-\alpha-1} \tag{5.7.72}$$
$$\mathbb{E}\, X = \frac{\alpha}{\alpha - 1} x_0, \quad \alpha > 1 \tag{5.7.73}$$

Properties:

- power law: $\log f_X(x) = log\left(\alpha\, x_0^{\alpha}\right) - (\alpha + 1) \log x$ 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.

Shifted Pareto, where the x values start from 0, is called Lomax Distribution or Pareto-II distribution.

5.7.15 Pööratud normaaljaotus

Tihedusfunktsioon:
$$f(t) = \frac{1}{t^{\frac{3}{2}}} \phi\left(\frac{\mu t - 1}{\sigma \sqrt{t}}\right) \tag{5.7.74}$$
ja jaotusfunktsioon:
$$F(t) = \Phi\left(\frac{\mu t - 1}{\sigma \sqrt{t}}\right) - e^{2\frac{\mu}{\sigma^2}} \Phi\left(-\frac{\mu t + 1}{\sigma \sqrt{t}}\right). \tag{5.7.75}$$
Kõik momendid on olemas kui $\mu > 0$:
$$\mathbb{E}\, T = \frac{1}{\mu} \tag{5.7.76}$$
$$\operatorname{Var} T = \frac{\sigma^2}{\mu^3}. \tag{5.7.77}$$
Kui $\mu = 0$, on jaotus korralik, positiivsed momendid aga puuduvad.

5.7.16 t-Distribution

Used for *t-test*. For n *degrees of freedom*:
$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{w}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \tag{5.7.78}$$
$$\mathbb{E}\, X = 0 \tag{5.7.79}$$
$$\operatorname{Var} X = \frac{n}{n-2} \tag{5.7.80}$$
$$\text{skewness } g_1 = 0 \tag{5.7.81}$$
$$\text{curtosis } g_2 = \frac{3n - 6}{n - 4} \qquad (n > 4) \tag{5.7.82}$$
Special case where $n = 1$ is Cauchy distribution.

5.7.17 Triangular Distribution

$$F(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \le x \le 1 \\ 1 & x > 1 \end{cases} \tag{5.7.83}$$

$$f(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases} \tag{5.7.84}$$

$$\mathbb{E}\, X = \frac{2}{3} \tag{5.7.85}$$

$$\mathbb{E}\, X^2 = \frac{1}{2} \tag{5.7.86}$$

$$\operatorname{Var} X = \frac{1}{18} \tag{5.7.87}$$

5.7.18 Type-1 Extreme Value Distribution EV_1

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

$$f(x) = \mathrm{e}^{-x}\mathrm{e}^{-\mathrm{e}^{-x}} \tag{5.7.88}$$

$$F(x) = \mathrm{e}^{-\mathrm{e}^{-x}}. \tag{5.7.89}$$

$$\mathbb{E}\, x = -c \approx 0,5772 \tag{5.7.90}$$

5.7.19 Type-2 Extreme Value Distribution EV_2

(Also Fréchet distribution or inverse Weibull distribution.)

$$f(x) = \alpha x^{-1-\alpha}\mathrm{e}^{-x^{-\alpha}} \tag{5.7.91}$$

$$F(x) = \mathrm{e}^{-x^{-\alpha}} \tag{5.7.92}$$

5.7.20 Uniform Distribution $Unif(a, b)$

$$\text{pdf} \qquad f(x) = \frac{1}{b-a}, \quad a \le x \le b \tag{5.7.93}$$

$$\text{MGF} \qquad M(s) = \frac{\mathrm{e}^{tb} - \mathrm{e}^{ta}}{t(b-a)} \tag{5.7.94}$$

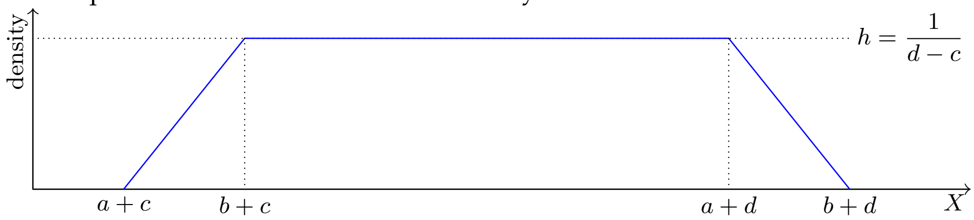
Moments:

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

$$X \sim Unif(a, b) \qquad Y \sim Unif(c, d) \tag{5.7.95}$$

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



5.7.21 Weibulli jaotus

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

$$F(t) = 1 - e^{-(\lambda t)^{\alpha}} \tag{5.7.96}$$

$$f(t) = \alpha\lambda^{\alpha} t^{\alpha-1} e^{-(\lambda t)^{\alpha}} \tag{5.7.97}$$

$$\theta(t) = \alpha\lambda^{\alpha} t^{\alpha-1} \tag{5.7.98}$$

where λ is scale- and α is the shape parameter.

5.8 Multivariate Continuous Distributions

5.8.1 Dirichlet distribution

Multivariate generalization of beta distribution.

$$\text{pdf} \qquad f(\boldsymbol{x};\boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})}\prod_{k=1}^{K} x_k^{\alpha_k-1}\mathbb{1}(\boldsymbol{x}\in S_k) \tag{5.8.1}$$

where S_K is the probability simplex:

$$S_K = \{\boldsymbol{x} : 0 \le x_k \le 1, \sum_{k=1}^{K} x_k = 1\} \tag{5.8.2}$$

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

$$B(\boldsymbol{\alpha}) = \frac{\prod_{k=1}^{K}\Gamma(\alpha_k)}{\Gamma(\alpha_0)} \qquad \alpha_0 = \sum_{k=1}^{K}\alpha_k \tag{5.8.3}$$

Mode:

$$m_k = \frac{\alpha_k - 1}{\alpha_0 - K} \tag{5.8.4}$$

Moments:

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

Properties $\boldsymbol{\alpha} = (1, 1, 1)$ gives uniform distribution.

5.8.2 Multivariate Normal $N(\mu, \Sigma)$

N-mõõtmelise normaajaotuse jaotusfunktsioon: Olgu

$$\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma), \tag{5.8.5}$$

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

$$f_X(\boldsymbol{x}) = (2\pi)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}\mathrm{e}^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}. \tag{5.8.6}$$

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

The exponent in the distribution function can be expressed as

$$-\frac{1}{2}\left(\frac{\sigma_2 x_1^2 - 2\sigma_{12}x_1x_2 + \sigma_1^2 x_2^2}{\sigma_1^2\sigma_2^2 - \sigma_{12}^2}\right) = -\frac{1}{2}\left[\frac{\left(x_1 - \frac{\sigma_{12}}{\sigma_2^2}x_2\right)^2}{\sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}} + \frac{x_2^2}{\sigma_2^2}\right]. \tag{5.8.7}$$

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:

$$(X_1|X_2 = x_2) \sim N\left(\frac{\sigma_{12}}{\sigma_2^2}x_2, \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}\right) \qquad X_2 \sim N(0, \sigma_2^2) \tag{5.8.8}$$

$$(X_2|X_1 = x_1) \sim N\left(\frac{\sigma_{12}}{\sigma_1^2}x_1, \sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}\right) \qquad X_1 \sim N(0, \sigma_1^2) \tag{5.8.9}$$

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

$$f_{X1|X2<a}(x_1) = \frac{1}{\sigma_1}\frac{\phi\left(\frac{x_1}{\sigma_1}\right)}{\Phi\left(\frac{a}{\sigma_2}\right)}\Phi\left(\frac{a - \frac{\sigma_{12}}{\sigma_1^2}x_1}{\sqrt{\sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}}}\right) \tag{5.8.10}$$

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

$$f_{X1|X2>a}(x_1) = \frac{1}{\sigma_1}\frac{\phi\left(\frac{x_1}{\sigma_1}\right)}{\Phi\left(-\frac{a}{\sigma_2}\right)}\Phi\left(-\frac{a - \frac{\sigma_{12}}{\sigma_1^2}x_1}{\sqrt{\sigma_2^2 - \frac{\sigma_{12}^2}{\sigma_1^2}}}\right) \tag{5.8.11}$$

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

$$(X_1|X_2 = x_2) \sim N\left(\mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(x_2 - \mu_2), \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2}\right) \tag{5.8.12}$$

(follows from 5.8.7) and

$$\mathbb{E}[X_1|X_2 < a] = \mu_1 - \frac{\sigma_{12}}{\sigma_2}\lambda\left(\frac{a - \mu_2}{\sigma_2}\right) \tag{5.8.13}$$

$$\mathbb{E}[X_1|X_2 > a] = \mu_1 + \frac{\sigma_{12}}{\sigma_2}\lambda\left(\frac{\mu_2 - a}{\sigma_2}\right) \tag{5.8.14}$$

$$\mathbb{E}[X_1^2|X_2 < a] = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^3}a\lambda\left(\frac{a}{\sigma_2}\right) \tag{5.8.15}$$

$$\mathbb{E}[X_1^2|X_2 > a] = \sigma_1^2 + \frac{\sigma_{12}^2}{\sigma_2^3}a\lambda\left(-\frac{a}{\sigma_2}\right) \tag{5.8.16}$$

$$\mathbb{E}[X_1^2|X_2 \in \mathcal{A}] = \frac{\sigma_{12}^2}{\sigma_2^4}\,\mathbb{E}[X_2^2|X_2 \in \mathcal{A}] + \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} \tag{5.8.17}$$

$$\mathrm{Var}\left[X_1|X_2 < a\right] = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^3}a\lambda\left(\frac{a}{\sigma_2}\right) - \frac{\sigma_{12}^2}{\sigma_2^2}\lambda^2\left(\frac{a}{\sigma_2}\right) \tag{5.8.18}$$

$$\mathrm{Var}\left[X_1|X_2 > a\right] = \sigma_1^2 + \frac{\sigma_{12}^2}{\sigma_2^3}a\lambda\left(-\frac{a}{\sigma_2}\right) - \frac{\sigma_{12}^2}{\sigma_2^2}\lambda^2\left(-\frac{a}{\sigma_2}\right) \tag{5.8.19}$$

Proof: write (5.8.12) $\Rightarrow X_1 = \mu_1 + \frac{\sigma_{12}}{\sigma_2}(X_2 - \mu_2) + E$, 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

$$\frac{1}{\sigma_1}\phi\left(\frac{x-ay}{\sigma_1}\right)\frac{1}{\sigma_2}\phi\left(\frac{y-b}{\sigma_2}\right) = \frac{1}{\sigma_x}\phi\left(\frac{x-ab}{\sigma_x}\right)\frac{1}{\sigma_y}\phi\left(\frac{y-\mu_y}{\sigma_y}\right), \tag{5.8.20}$$

where

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

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

The same in multi-dimensional case:

$$\begin{aligned}
\frac{1}{\sigma_1}\phi\left(\frac{x_1-y}{\sigma_1}\right)\frac{1}{\sigma_1}\phi\left(\frac{x_2-y}{\sigma_1}\right)\dots\frac{1}{\sigma_1}\phi\left(\frac{x_n-y}{\sigma_1}\right)\frac{1}{\sigma_2}\phi\left(\frac{y}{\sigma_2}\right) &= \\
= \prod_{i=1}^{n}\frac{1}{\sigma_1}\phi\left(\frac{x_i-y}{\sigma_1}\right)\frac{1}{\sigma_2}\phi\left(\frac{y}{\sigma_2}\right) &= \\
= \frac{1}{\sigma_x}\phi\left(\frac{x_1}{\sigma_x}\right)\frac{1}{\sigma_x}\phi\left(\frac{x_2}{\sigma_x}\right)\dots\frac{1}{\sigma_x}\phi\left(\frac{x_n}{\sigma_x}\right)\frac{1}{\sigma_y}\phi\left(\frac{y-\mu_y}{\sigma_y}\right) &= \\
= \prod_{i=1}^{n}\frac{1}{\sigma_x}\phi\left(\frac{x_i}{\sigma_x}\right)\frac{1}{\sigma_y}\phi\left(\frac{y-\mu_y}{\sigma_y}\right)
\end{aligned} \tag{5.8.21}$$

where

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

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

5.8.3 Wishart distribution

Wishart distribution is a generalization of gamma distribution to positive definite matrices.

Probability density:

$$f(\mathsf{X}) = \frac{1}{Z_{Wi}}|\mathsf{X}|^{\frac{1}{2}(\nu-D-1)}\mathrm{e}^{-\frac{1}{2}\operatorname{Tr}(\mathsf{X}\mathsf{S}^{-1})} \tag{5.8.22}$$

where ν is degrees of freedom, S is scale matrix, and

$$Z_{Wi} = 2^{\frac{1}{2}\nu D}\Gamma_D\left(\frac{\nu}{2}\right)|\mathsf{S}|^{\frac{\nu}{2}}. \tag{5.8.23}$$

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

Properties:

$$\mathbb{E}\,\mathsf{X} = \nu\mathsf{S} \tag{5.8.24}$$

$$\text{mode} = (\nu - D - 1)\mathsf{S} \quad \text{for} \quad \nu > D + 1 \tag{5.8.25}$$

5.9 Distribution Families

5.9.1 Exponential Family

Exponential family is a distribution whose density can be written as

$$p(x|\boldsymbol{\eta}) = h(x)\mathrm{e}^{\eta' T(x) - A(\eta)} = \frac{h(x)}{Z(\boldsymbol{\eta})}\mathrm{e}^{\eta' T(x)}, \tag{5.9.1}$$

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

$$p(x|\boldsymbol{\theta}) = h(x)\mathrm{e}^{\boldsymbol{\eta}(\boldsymbol{\theta})' T(x) - A(\eta(\boldsymbol{\theta}))}. \tag{5.9.2}$$

5.9.2 Stabiilne pere

Mittenegatiivsesse stabiilsesse perre kuuluvad jaotused, mille momendifunktsioon on

$$M_x(s) = e^{-s^{\alpha}}, \qquad 0 < \alpha \leq 1. \tag{5.9.3}$$

Stabiilsel pere omadused:

1. kui juhusliku muutuja X_i jaotusfunktsioon on G_α mis kuulub stabiilsesse perre, siis juhusliku muutuja

$$Y = n^{-\frac{1}{\alpha}} \sum_{i=1}^{N} X_i \tag{5.9.4}$$

   jaotusfunktsioon on kah G_α.

2. Momendifunktsiooni tuletis

$$M'_x(s) = -\alpha s^{\alpha-1} M(s) \tag{5.9.5}$$

5.10 Functions of Random Variables

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_K)$ be a vector of i.i.d random variables with $\mathbb{E}\, X_i X_j = \delta_{ij}\sigma^2$, and A be a $K \times K$ matrix. Then

$$\mathbb{E}[\boldsymbol{X}^{\intercal}\,\mathsf{A}\,\boldsymbol{X}] = \operatorname{Tr}\mathsf{A} \cdot \mathbb{E}\, X^2 \tag{5.10.1}$$

6 Estimators

6.1 M-Estimators

6.1.1 Variance

Let an estimator solve

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

From Taylor approximation

$$\sum_i h_i(\hat{\boldsymbol{\theta}}) = \sum_i h_i(\boldsymbol{\theta}_0) + \frac{\partial}{\partial \boldsymbol{\theta}} \sum_i h_i(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}_0} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = 0 \tag{6.1.2}$$

from where

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = -\left(\frac{\partial}{\partial \boldsymbol{\theta}} \sum_i h_i(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}_0} \right)^{-1} \sum_i h_i(\boldsymbol{\theta}_0). \tag{6.1.3}$$

The estimate for variance is

$$\operatorname{Var} \hat{\boldsymbol{\theta}} = \hat{\mathsf{A}}^{-1} \widehat{\operatorname{Var} H} \hat{\mathsf{A}}^{-1} \tag{6.1.4}$$

where

$$\hat{\mathsf{A}} = \frac{\partial}{\partial \boldsymbol{\theta}} \sum_i h_i(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}_0} \tag{6.1.5}$$

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

6.2 Maximum likelihood

6.2.1 Definition

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

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

The *score* is defined as

$$g(\boldsymbol{\vartheta}|x_1, x_2, \ldots, x_n) = \frac{\partial}{\partial \boldsymbol{\theta}} \ell(\boldsymbol{\vartheta}|x_1, x_2, \ldots, x_n) \tag{6.2.2}$$

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

$$\hat{\boldsymbol{\vartheta}} = \arg\max_{\boldsymbol{\vartheta}} \ell(\boldsymbol{\vartheta}|x_1, x_2, \ldots, x_n). \tag{6.2.3}$$

6.2.2 Information matrix

Information matrix is defined as

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

(information equality.)

6.2.3 Relationship to Kullback-Leibler divergence

ML estimator can be written as

$$\hat{\boldsymbol{\vartheta}} = \arg\max_{\boldsymbol{\vartheta}} \int \log f(x|\boldsymbol{\vartheta})\, \mathrm{d}F_n(x), \tag{6.2.5}$$

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

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

Further, we may write the estimator as

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

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

$$\hat{\boldsymbol{\vartheta}} = \arg\min_{\boldsymbol{\vartheta}} \int \log \frac{f(x|\boldsymbol{\vartheta}_0)}{f(x|\boldsymbol{\vartheta})}\, \mathrm{d}F_n(x) = \arg\min_{\boldsymbol{\vartheta}} KL(f|\boldsymbol{\vartheta}_0\|f|\boldsymbol{\vartheta}), \tag{6.2.8}$$

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

$$V_{GMM} = \frac{1}{n}\left[\mathsf{\Gamma}'\,\mathsf{W}\mathsf{\Gamma}\right]^{-1} \tag{6.3.1}$$

where W is the weighting matrix and

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

Optimal Weighting Matrix

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

6.3.1 Optimal Weighting Matrices and Asymptotic Variances

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

For the moment condition

$$\mathbb{E}\,X - \mu = 0 \tag{6.3.4}$$

the optimal weighting matrix:

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

For the moment condition

$$\mathbb{E}\,X^2 - \mu_2 = 0 \tag{6.3.6}$$

the optimal weighting matrix:

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

For the moment condition

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

Matrix

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

we have the optimal weighting matrix:

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

and the variance

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

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

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

6.4 Entropy Distance

6.4.1 Entropy Distance

Lin (1991) defines *entropy distance:*

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

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

$$u_t = \varrho u_{t-1} + \varepsilon_t \tag{7.1.1}$$

ning ε_t väärtused eri ajaperioodidel on sõltumatud. Et protsess oleks stabiilne peab ϱ 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

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

ning ε_t väärtused eri ajaperioodidel on sõltumatud.

7.2 Hulkumine

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

$$z_{t+1} = z_t + \varepsilon_{t+1}. \tag{7.2.1}$$

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 ületa barjääri α on

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

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

8 Dynamic Models

8.1 Contagion in Random Network

Look at population of size N where n are infected. The individuals are meeting at random with Poisson rate λ. $p = n/N$ is the probability a random individual in the population is infected. The dynamics is governed by

$$\mathrm{d}n = \lambda n(1 - p)\,\mathrm{d}t \tag{8.1.1}$$

n infected persons meet someone at rate λ. With probability $1 - p$ the one they meet was uninfected and will be infected.

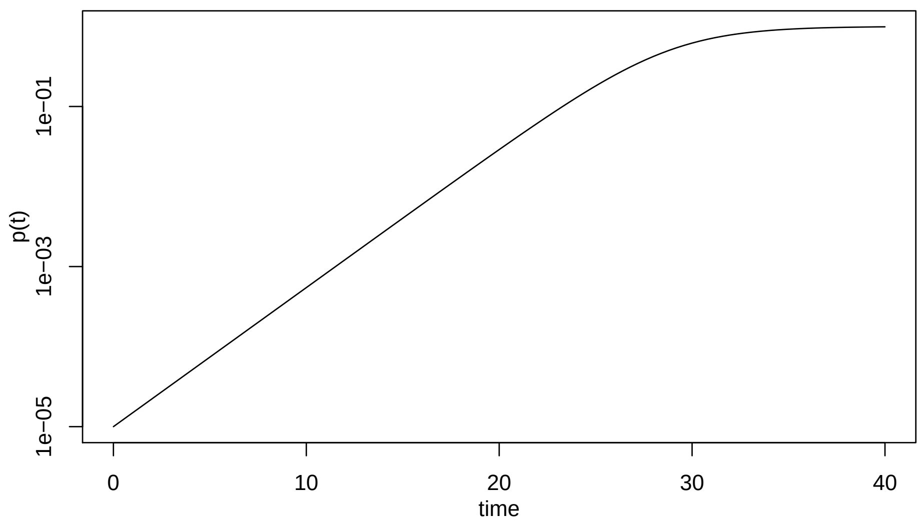
Divide both sides by N to receive

$$\mathrm{d}p = \lambda p(1 - p)\,\mathrm{d}t \quad \text{or} \quad \frac{\mathrm{d}p}{p(1 - p)} = \lambda\,\mathrm{d}t \tag{8.1.2}$$

Assume initial state at $t = 0$ is p_0 infection probability. The solution is

$$p(t) = \frac{p_0}{p_0 + (1 - p_0)\mathrm{e}^{-\lambda t}}. \tag{8.1.3}$$

The initial exponential growth with rate λ flattens out later:



9 Statistical Models

9.1 Linear Regression

Loss function (MSE):

$$L(\boldsymbol{\beta}) = \frac{1}{N}(\boldsymbol{y} - \mathsf{X}\boldsymbol{\beta})^{\top}(\boldsymbol{y} - \mathsf{X}\boldsymbol{\beta}) \tag{9.1.1}$$

Its gradient:

$$\frac{\partial}{\partial \boldsymbol{\beta}} L(\boldsymbol{\beta}) = -\frac{2}{N}\mathsf{X}^{\top}(\boldsymbol{y} - \mathsf{X}\boldsymbol{\beta}) \qquad \text{matrix form} \tag{9.1.2}$$

$$= -\frac{2}{N}\sum_{i=1}^{N}(\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta}) \cdot \boldsymbol{x}_i \qquad \text{vector form} \tag{9.1.3}$$

Predictions:

$$\hat{\boldsymbol{y}} = \mathsf{P}\,\boldsymbol{y} = \mathsf{X}\,(\mathsf{X}^{\top}\mathsf{X})^{-1}\,\mathsf{X}^{\top}\,\boldsymbol{y} \tag{9.1.4}$$

where $\mathsf{P} = \mathsf{X}\,(\mathsf{X}^{\top}\mathsf{X})^{-1}\,\mathsf{X}^{\top}$ is the *projection matrix*.

$$\text{Residuals} \quad \hat{\boldsymbol{y}} = \mathsf{M}\,\boldsymbol{y} = (\mathsf{I} - \mathsf{P})\,\boldsymbol{y} \tag{9.1.5}$$

$$\text{MSE} \quad \hat{\boldsymbol{y}} = \frac{1}{N}\,\boldsymbol{y}^{\top}\,\mathsf{M}^{\top}\mathsf{M}\,\boldsymbol{y} = \frac{1}{N}\,\boldsymbol{y}^{\top}\,\mathsf{M}^{2}\,\boldsymbol{y} \tag{9.1.6}$$

Gradient in matrix

9.2 Tobit-2 model

Definition:

$$y_{1i}^{*} = \boldsymbol{z}_i'\boldsymbol{\gamma} + u_{1i} \tag{9.2.1}$$

$$y_{2i}^{*} = \boldsymbol{x}_i'\boldsymbol{\beta} + u_{2i} \tag{9.2.2}$$

$$y_{1i} = \begin{cases} 1, & \text{if} \quad y_{1i}^{*} > 0 \\ 0, & \quad y_{1i}^{*} \leq 0. \end{cases} \tag{9.2.3}$$

$$y_{2i} = \begin{cases} y_{2i}^{*}, & \text{if} \quad y_{1i}^{*} > 0 \\ 0, & \quad y_{1i}^{*} \leq 0. \end{cases} \tag{9.2.4}$$

Assume

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \varrho \\ \varrho & \sigma^2 \end{pmatrix}\right). \tag{9.2.5}$$

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

$$\mathbb{E}[Y_2|Y_1 > 0, \boldsymbol{x}, \boldsymbol{z}] = \boldsymbol{x}'\beta + \mathbb{E}[U_2|U_1 > -\boldsymbol{z}'\gamma] = \boldsymbol{x}'\beta + \varrho\sigma\lambda(-\boldsymbol{z}'\gamma) \tag{9.2.6}$$

$$\operatorname{Var}[Y_2|Y_1 > 0, \boldsymbol{x}, \boldsymbol{z}] = \mathbb{E}[U_2|U_1 > -\boldsymbol{z}'\gamma] = \sigma^2 + \varrho^2\sigma^2[-\boldsymbol{z}'\gamma\lambda(\boldsymbol{z}'\gamma) - \lambda^2(\boldsymbol{z}'\gamma)] \tag{9.2.7}$$

where $\lambda(x) = \phi(x)/\Phi(x)$; $\Phi(\cdot)$ and $\phi(\cdot)$ are the normal cumulative distribution function and density function respectively. ϱ and σ can be estimated regressing

y_{2i} on $\boldsymbol{x}_i$ and $\lambda(-\boldsymbol{z}'\gamma)$. From the coefficient of the latter, β_λ and the residual variance s^2, one can isolate ϱ and σ:

$$\hat{\sigma}^2 = s^2 + \beta_\lambda^2[\lambda^2(\boldsymbol{z}'\gamma) - \boldsymbol{z}'\gamma\lambda(\boldsymbol{z}'\gamma)] \tag{9.2.8}$$

$$\hat{\varrho} = \frac{\beta_\lambda}{\hat{\sigma}}. \tag{9.2.9}$$

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

Denote:

$$r = \sqrt{1 - \varrho^2} \tag{9.2.10}$$

$$u_{2i} = y_{2i} - \boldsymbol{x}_i'\boldsymbol{\beta} \tag{9.2.11}$$

$$B_i = \frac{\boldsymbol{z}_i'\boldsymbol{\gamma} + \frac{\varrho}{\sigma}u_{2i}}{r} \tag{9.2.12}$$

$$C(B) = -\frac{\Phi(B)\phi(B)B + \phi(B)^2}{\Phi(B)^2} \tag{9.2.13}$$

The contribution of observation i to the log-likelihood:

$$\ell = \sum_{i:y_{1i}\le 0} \log\Phi(-\boldsymbol{z}_i'\boldsymbol{\gamma}) + \tag{9.2.14}$$

$$+ \sum_{i:y_{1i}>0}\left[\log\Phi(B_i) - \frac{1}{2}\log 2\pi - \log\sigma - \frac{1}{2}\frac{u_{2i}^2}{\sigma^2}\right]. \tag{9.2.15}$$

The gradient of the log-likelihood is:

$$\frac{\partial\ell}{\partial\boldsymbol{\gamma}} = \sum_{i:y_{1i}\le 0} -\lambda(\boldsymbol{z}_i'\boldsymbol{\gamma})\boldsymbol{z}_i + \sum_{i:y_{1i}>0}\lambda(B_i)\frac{\boldsymbol{z}_i}{r} \tag{9.2.16}$$

$$\frac{\partial\ell}{\partial\boldsymbol{\beta}} = \sum_{i:y_{1i}>0}\left[\frac{u_{2i}}{\sigma^2} - \lambda(B_i)\frac{\varrho}{\sigma}\frac{1}{r}\right]\boldsymbol{x}_i \tag{9.2.17}$$

$$\frac{\partial l}{\partial\sigma} = \sum_{i:y_{1i}>0}\left[\frac{u_{2i}^2}{\sigma^3} - \frac{1}{\sigma} - \lambda(B_i)\frac{\varrho}{\sigma^2}\frac{u_{2i}}{r}\right] \tag{9.2.18}$$

$$\frac{\partial\ell}{\partial\varrho} = \sum_{i:y_{1i}>0}\lambda(B_i)\frac{\frac{1}{\sigma}u_{2i} + \varrho\boldsymbol{z}_i'\boldsymbol{\gamma}}{r^3}. \tag{9.2.19}$$

Hessian components are

$$\frac{\partial^2\ell}{\partial\boldsymbol{\gamma}\boldsymbol{\gamma}'} = -\sum_{i:y_{1i}=0} C(-\boldsymbol{z}_i'\boldsymbol{\gamma})\boldsymbol{z}_i\boldsymbol{z}_i' + \sum_{i:y_{1i}=1}\frac{C(B)}{r}\boldsymbol{z}_i\boldsymbol{z}_i' \tag{9.2.20}$$

$$\frac{\partial^2\ell}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\beta}'} = -\sum_{i:y_{1i}=1} C(B)\frac{1}{\sigma}\frac{\varrho}{r}\boldsymbol{z}_i\boldsymbol{x}_i' \tag{9.2.21}$$

$$\frac{\partial^2\ell}{\partial\boldsymbol{\gamma}\partial\sigma} = -\sum_{i:y_{1i}=1} C(B)\frac{\varrho u_2}{\sigma^2 r^2}\boldsymbol{z}_i \tag{9.2.22}$$

$$\frac{\partial^2 \ell}{\partial \gamma \partial \varrho} = \sum_{i:y_{1i}=1} \left[C(B) \frac{\frac{u2}{\sigma} + \varrho z_i' \gamma}{r^4} + \lambda(B) \frac{\varrho}{r^3} \right] z_i \tag{9.2.23}$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta'} = \sum_{i:y_{1i}=1} \frac{1}{\sigma^2} \left[\frac{\varrho^2}{r^2} C(B) - 1 \right] x_i x_i' \tag{9.2.24}$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \sigma} = \sum_{i:y_{1i}=1} \left[C(B) \frac{\varrho^2}{\sigma^3} \frac{u_2}{r^2} + \frac{\varrho}{\sigma^2} \frac{\lambda(B)}{r} - 2 \frac{u_2}{\sigma^3} \right] x_i \tag{9.2.25}$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \varrho} = \sum_{i:y_{1i}=1} \left[-C(B) \frac{\frac{u_2}{\sigma} + \varrho z_i' \gamma}{r^4} \frac{\varrho}{\sigma} - \frac{\lambda(B)}{\sigma} \frac{1}{r^3} \right] x_i \tag{9.2.26}$$

$$\frac{\partial^2 \ell}{\partial \sigma_3^2} = \sum_{i:y_{1i}=1} \left[\frac{1}{\sigma^2} - 3 \frac{u_2^2}{\sigma^4} + 2\lambda(B) \frac{u2}{r} \frac{\varrho}{\sigma^3} + \frac{\varrho^2}{\sigma^4} \frac{u_2^2}{r^2} C(B) \right] \tag{9.2.27}$$

$$\frac{\partial^2 \ell}{\partial \sigma \partial \varrho} = -\frac{1}{r^3} \sum_{i:y_{1i}=1} \frac{u_2}{\sigma^2} \left[C(B) \frac{\varrho \left(\frac{u_2}{\sigma} + \varrho z_i' \gamma \right)}{r} + \lambda(B) \right] \tag{9.2.28}$$

$$\frac{\partial^2 \ell}{\partial \varrho^2} = \sum_{i:y_{1i}=1} \left[C(B) \left(\frac{\frac{u_2}{\sigma} + \varrho z_i' \gamma}{r^3} \right)^2 + \lambda(B) \frac{z_i' \gamma (1 + 2\varrho^2) + 3\varrho \frac{u_2}{\sigma}}{r^5} \right] \tag{9.2.29}$$

9.3 Tobit-2 Model with Binary Outcome

The underlying latent model:

$$y_i^{S*} = \beta^{S\prime} x_i^S + \epsilon_i^S \tag{9.3.1}$$

$$y_i^{O*} = \beta^{O\prime} x_i^O + \epsilon_i^0 \tag{9.3.2}$$

$$y_i^S = \begin{cases} 1, & \text{if} \quad y_i^{S*} > 0 \\ 0, & y_i^{S*} \le 0. \end{cases} \tag{9.3.3}$$

$$y_i^O = \begin{cases} \text{undetermined}, & \text{if} \quad y_i^S = 0 \quad \text{(case 1)} \\ 0, & \text{if} \quad y_i^{O*} \le 0 \quad \text{and} \quad y_i^S = 1 \quad \text{(case 2)} \\ 1, & \text{if} \quad y_i^{O*} > 0 \quad \text{and} \quad y_i^S = 1 \quad \text{(case 3)} \end{cases} \tag{9.3.4}$$

Assume

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix} \right). \tag{9.3.5}$$

The log-likelihood function contains 3 components, corresponding to the cases in (9.3.4):

$$\ell = \sum_{i \in \text{case 1}} \log \Phi\left(-\beta^{S\prime} x_i^S \right) \tag{9.3.6}$$

$$+ \sum_{i \in \text{case 2}} \log \left[1 - \Phi\left(-\beta^{S\prime} x_i^S \right) - \bar{\Phi}_2 \left(\begin{pmatrix} -\beta^{S\prime} x_i^S \\ -\beta^{O\prime} x_i^O \end{pmatrix}, \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix} \right) \right] \tag{9.3.7}$$

$$+\sum_{i\in\text{case 3}}\log\bar{\Phi}_2\left(\begin{pmatrix}-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\\-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O\end{pmatrix},\begin{pmatrix}1&\varrho\\\varrho&1\end{pmatrix}\right),\tag{9.3.8}$$

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:

$$\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\beta}^S}\ell =& \sum_{i\in\text{case 1}}\frac{1}{\mathcal{L}_i}\phi\left(-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\boldsymbol{x}_i^S\\
&+\sum_{i\in\text{case 2}}\frac{1}{\mathcal{L}_i}\phi\left(\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\bar{\Phi}\left(\frac{\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O-\varrho\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S}{\sqrt{1-\varrho^2}}\right)\boldsymbol{x}_i^S\\
&+\sum_{i\in\text{case 3}}\frac{1}{\mathcal{L}_i}\phi\left(\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\right)\Phi\left(\frac{\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O-\varrho\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S}{\sqrt{1-\varrho^2}}\right)\boldsymbol{x}_i^S\\
\frac{\partial}{\partial\boldsymbol{\beta}^O}\ell =& \sum_{i\in\text{case 2}}\frac{1}{\mathcal{L}_i}\phi\left(\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O\right)\bar{\Phi}\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S-\varrho\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sqrt{1-\varrho^2}}\right)\boldsymbol{x}_i^O\\
&+\sum_{i\in\text{case 3}}\frac{1}{\mathcal{L}_i}\phi\left(\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O\right)\Phi\left(\frac{\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S-\varrho\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O}{\sqrt{1-\varrho^2}}\right)\boldsymbol{x}_i^O\\
\frac{\partial}{\partial\varrho}\ell =& -\sum_{i\in\text{case 2}}\phi_2\left(\begin{pmatrix}-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\\-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O\end{pmatrix},\begin{pmatrix}1&\varrho\\\varrho&1\end{pmatrix}\right)+\sum_{i\in\text{case 3}}\phi_2\left(\begin{pmatrix}-\boldsymbol{\beta}^{S\prime}\boldsymbol{x}_i^S\\-\boldsymbol{\beta}^{O\prime}\boldsymbol{x}_i^O\end{pmatrix},\begin{pmatrix}1&\varrho\\\varrho&1\end{pmatrix}\right),
\end{aligned}$$

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

9.4 Tobit-5 Model

Definitsioon (lühiduse mõttes on indeks i ära jäetud):

$$y_1^* = \boldsymbol{Z}'\boldsymbol{\gamma} + u_1 \tag{9.4.1}$$
$$y_2^* = \boldsymbol{X}'\boldsymbol{\beta_2} + u_2 \tag{9.4.2}$$
$$y_3^* = \boldsymbol{X}'\boldsymbol{\beta_3} + u_3 \tag{9.4.3}$$
$$y_2 = \begin{cases} y_2^* & \text{kui} \quad y_1^* \le 0 \\ 0 & \phantom{\text{kui}} \quad y_1^* > 0 \end{cases} \tag{9.4.4}$$
$$y_3 = \begin{cases} y_3^* & \text{kui} \quad y_1^* > 0 \\ 0 & \phantom{\text{kui}} \quad y_1^* \le 0 \end{cases} \tag{9.4.5}$$

Eeldatakse et jääkliikmete jaotus on niisugune:

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \varrho_2\sigma_2 & \varrho_3\sigma_3 \\ \varrho_2\sigma_2 & \sigma_2^2 & \sigma_{23} \\ \varrho_3\sigma_3 & \sigma_{23} & \sigma_3^2 \end{pmatrix}\right). \tag{9.4.6}$$

9.4.1 Heckmani kahesammuline hinnang

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

$$\begin{aligned} y_2 &= \boldsymbol{X}'\boldsymbol{\beta_2} - \varrho_2\sigma_2\lambda(-\boldsymbol{Z}'\boldsymbol{\gamma}) + e_2 \\ y_3 &= \boldsymbol{X}'\boldsymbol{\beta_3} + \varrho_3\sigma_3\lambda(\boldsymbol{Z}'\boldsymbol{\gamma}) + e_3 \end{aligned} \tag{9.4.7}$$

Kusjuures

$$\begin{aligned} \sigma_{e2} &= \sigma_2^2\left\{1 - \varrho_2^2\left[\lambda^2(-\boldsymbol{Z}'\boldsymbol{\gamma}) - \boldsymbol{Z}'\boldsymbol{\gamma}\lambda(-\boldsymbol{Z}'\boldsymbol{\gamma})\right]\right\} \\ \sigma_{e3} &= \sigma_3^2\left\{1 - \varrho_3^2\left[\lambda^2(\boldsymbol{Z}'\boldsymbol{\gamma}) + \boldsymbol{Z}'\boldsymbol{\gamma}\lambda(-\boldsymbol{Z}'\boldsymbol{\gamma})\right]\right\} \end{aligned} \tag{9.4.8}$$

Kui lähendada seost (9.4.7) OLS-ga, siis saab λ koefitsendi ja dispersiooni hinnangu abil leida $\hat{\varrho}$ ja $\hat{\sigma}$. Märkus: $\hat{\varrho}$ ei pruugi olla -1 ja 1 vahel.

9.4.2 Maksimum-laiklikhuud hinnang

Mudeli log-laiklihuud on:

$$\begin{aligned} l &= -\frac{N}{2}\log 2\pi + \\ &+ \sum_{i\in\text{case 2}}\left\{-\log\sigma_2 - \frac{1}{2}\left(\frac{u_2}{\sigma_2}\right)^2 + \log\Phi\left[-\frac{\boldsymbol{Z}'\boldsymbol{\gamma} + \frac{\varrho_2}{\sigma_2}\left(y_2 - \boldsymbol{X}_i'\boldsymbol{\beta}_2\right)}{\sqrt{1-\varrho_2^2}}\right]\right\} \\ &+ \sum_{i\in\text{case 3}}\left\{-\log\sigma_3 - \frac{1}{2}\left(\frac{y_3 - \boldsymbol{X}\boldsymbol{\beta}_3}{\sigma_3}\right)^2 + \log\Phi\left[\frac{\boldsymbol{Z}'\boldsymbol{\gamma} + \frac{\varrho_3}{\sigma_3}\left(y_3 - \boldsymbol{X}_i'\boldsymbol{\beta}_3\right)}{\sqrt{1-\varrho_3^2}}\right]\right\}. \end{aligned} \tag{9.4.9}$$

Valikud 2 ja 3 erinevad ainult avaldise märgi poolest funktsiooni Φ sees. Tuletised on:

$$\frac{\partial l}{\partial \gamma} = -\sum_2 \frac{\phi(B_2)}{\Phi(B_2)} \frac{\boldsymbol{Z}}{\sqrt{1-\varrho_2^2}} + \sum_3 \frac{\phi(B_3)}{\Phi(B_3)} \frac{\boldsymbol{Z}}{\sqrt{1-\varrho_3^2}} \tag{9.4.10}$$

$$\frac{\partial l}{\partial \boldsymbol{\beta}_2} = \sum_2 \left[\frac{\phi(B_2)}{\Phi(B_2)} \left(\frac{\varrho_2}{\sigma_2} \frac{\boldsymbol{X}}{\sqrt{1-\varrho_2^2}} \right) + \frac{u_2}{\sigma_2^2}\boldsymbol{X} \right] \tag{9.4.11}$$

$$\frac{\partial l}{\partial \sigma_2} = \sum_2 \left[-\frac{1}{\sigma_2} + \frac{\left(y_2 - \boldsymbol{X}'\boldsymbol{\beta}_2\right)^2}{\sigma_2^3} + \frac{\phi(B_2)}{\Phi(B_2)} \frac{\varrho_2}{\sigma_2^2} \frac{y_2 - \boldsymbol{X}'\boldsymbol{\beta}_2}{\sqrt{1-\varrho_2^2}} \right] \tag{9.4.12}$$

$$\frac{\partial l}{\partial \varrho_2} = -\sum_2 \frac{\phi(B_2)}{\Phi(B_2)} \frac{\frac{1}{\sigma_2}(y_2 - \boldsymbol{X}'\boldsymbol{\beta}_2) + \varrho_2 \boldsymbol{Z}'\boldsymbol{\gamma}}{(1-\varrho_2^2)^{\frac{3}{2}}} \tag{9.4.13}$$

$$\frac{\partial l}{\partial \boldsymbol{\beta}_3} = \sum_3 \left[-\frac{\phi(B_3)}{\Phi(B_3)} \left(\frac{\varrho_3}{\sigma_3} \frac{\boldsymbol{X}}{\sqrt{1-\varrho_3^2}} \right) + \frac{u_3}{\sigma_3^2}\boldsymbol{X} \right] \tag{9.4.14}$$

$$\frac{\partial l}{\partial \sigma_3} = \sum_3 \left[-\frac{1}{\sigma_3} + \frac{\left(y_3 - \boldsymbol{X}'\boldsymbol{\beta}_3\right)^2}{\sigma_3^3} - \frac{\phi(B_3)}{\Phi(B_3)} \frac{\varrho_3}{\sigma_3^2} \frac{y_3 - \boldsymbol{X}'\boldsymbol{\beta}_3}{\sqrt{1-\varrho_3^2}} \right] \tag{9.4.15}$$

$$\frac{\partial l}{\partial \varrho_3} = \sum_3 \frac{\phi(B_3)}{\Phi(B_3)} \frac{\frac{1}{\sigma_3}(y_3 - \boldsymbol{X}'\boldsymbol{\beta}_3) + \varrho_3 \boldsymbol{Z}'\boldsymbol{\gamma}}{(1-\varrho_3^2)^{\frac{3}{2}}} \tag{9.4.16}$$

Teised tuletised:

$$\frac{\partial^2 l}{\partial \boldsymbol{\gamma}^2} = \sum_2 \frac{C(B_2)}{1-\varrho_2^2} z_i z_i' + \sum_3 \frac{C(B_3)}{1-\varrho_3^2} z_i z_i' \tag{9.4.17}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}_2'} = -\sum_2 C(B_2) \frac{1}{\sigma_2} \frac{\varrho_2}{1-\varrho_2^2} \boldsymbol{Z}\boldsymbol{X}' \tag{9.4.18}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\gamma} \partial \sigma_2} = -\sum_2 \frac{\varrho_2 u_2}{\sigma_2^2 (1-\varrho_2^2)} C(B_2)\boldsymbol{Z} \tag{9.4.19}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\gamma} \partial \varrho_2} = \sum_2 \left[C(B_2) \frac{\frac{u_2}{\sigma_2} \varrho_2 \boldsymbol{Z}'\boldsymbol{\gamma}}{(1-\varrho_2^2)^2} - \lambda(B_2) \frac{\varrho_2}{(1-\varrho_2^2)^{\frac{3}{2}}} \right] \boldsymbol{Z} \tag{9.4.20}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\beta}_3'} = -\sum_3 C(B_3) \frac{1}{\sigma_3} \frac{\varrho_3}{1-\varrho_3^2} \boldsymbol{Z}\boldsymbol{X}' \tag{9.4.21}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\gamma} \partial \sigma_3} = -\sum_3 \frac{\varrho_3 u_3}{\sigma_3^2 (1-\varrho_3^2)} C(B_3)\boldsymbol{Z} \tag{9.4.22}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\gamma} \partial \varrho_3} = \sum_3 \left[C(B_3) \frac{\frac{u_3}{\sigma_3} \varrho_3 \boldsymbol{Z}' \boldsymbol{\gamma}}{(1-\varrho_3^2)^2} + \lambda(B_3) \frac{\varrho_3}{(1-\varrho_3^2)^{\frac{3}{2}}} \right] \boldsymbol{Z} \tag{9.4.23}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2'} = \sum_2 \frac{1}{\sigma_2^2} \left[\frac{\varrho_2^2}{1-\varrho_2^2} C(B_2) - 1 \right] \boldsymbol{X}\boldsymbol{X}' \tag{9.4.24}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta}_2 \partial \sigma_2} = \sum_2 \left[C(B_2) \frac{u_2}{\sigma_2^3} \frac{\varrho_2^2}{1-\varrho_2^2} - \frac{\lambda(B_2)}{\sigma_2^2} \frac{\varrho_2}{\sqrt{1-\varrho_2^2}} - 2 \frac{u_2}{\sigma_2^3} \right] \boldsymbol{X} \tag{9.4.25}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta}_2 \partial \varrho_2} = \sum_2 \left[-C(B_2) \frac{\frac{u_2}{\sigma_2} + \varrho_2 \boldsymbol{Z}' \boldsymbol{\gamma}}{(1-\varrho_2^2)^2} \frac{\varrho_2}{\sigma_2} + \frac{\lambda(B_2)}{\sigma_2} \frac{1}{(1-\varrho_2^2)^{\frac{3}{2}}} \right] \boldsymbol{X} \tag{9.4.26}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_3} = 0 \tag{9.4.27}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta}_2 \partial \sigma_3} = 0 \tag{9.4.28}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta}_2 \partial \varrho_3} = 0 \tag{9.4.29}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \sigma_2^2} &= \sum_2 \left[\frac{1}{\sigma_2^2} - 3 \frac{u_2^2}{\sigma_2^4} + \frac{u_2}{\sigma_2^4} \frac{\varrho_2^2}{1-\varrho_2^2} C(B_2) \right] - \\ &- 2 \sum_2 \lambda(B_2) \frac{u_2}{\sigma_2^3} \frac{\varrho_2}{\sqrt{1-\varrho_2^2}} \end{aligned} \tag{9.4.30}$$

$$\frac{\partial^2 l}{\partial \sigma_2 \partial \varrho_2} = \frac{1}{(1-\varrho_2^2)^{\frac{3}{2}}} \sum_2 \frac{u_2}{\sigma_2^2} \left[-C(B_2) \frac{\varrho_2 \left(\frac{u_2}{\sigma_2} + \varrho_2 \boldsymbol{Z}' \boldsymbol{\gamma} \right)}{\sqrt{1-\varrho_2^2}} + \lambda(B_2) \right] \tag{9.4.31}$$

$$\frac{\partial^2 l}{\partial \sigma_2 \partial \boldsymbol{\beta}_3} = 0 \tag{9.4.32}$$

$$\frac{\partial^2 l}{\partial \sigma_2 \partial \sigma_3} = 0 \tag{9.4.33}$$

$$\frac{\partial^2 l}{\partial \sigma_2 \partial \varrho_3} = 0 \tag{9.4.34}$$

$$\begin{aligned} \frac{\partial^2 l}{\partial \varrho_2^2} &= \sum_2 C(B_2) \left[\frac{\frac{u_2}{\sigma_2} + \varrho_2 \boldsymbol{Z}' \boldsymbol{\gamma}}{(1-\varrho_2^2)^{\frac{3}{2}}} \right]^2 - \\ &\quad - \sum_2 \frac{\phi(B_2)}{\Phi(B_2)} \frac{\boldsymbol{Z}' \boldsymbol{\gamma} (1 + 2\varrho_2^2) + 3 \varrho_2 \frac{u_2}{\sigma_2}}{(1-\varrho_2^2)^{\frac{5}{2}}} \end{aligned} \tag{9.4.35}$$

$$\frac{\partial^2 l}{\partial \varrho_2 \partial \boldsymbol{\beta}_3} = 0 \tag{9.4.36}$$

$$\frac{\partial^2 l}{\partial \varrho_2 \partial \sigma_3} = 0 \tag{9.4.37}$$

$$\frac{\partial^2 l}{\partial \varrho_2 \partial \varrho_3} = 0 \tag{9.4.38}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta}_3 \partial \boldsymbol{\beta}_3'} = \sum_3 \frac{1}{\sigma_3^2}\left[\frac{\varrho_3^2}{1-\varrho_3^2}C(B_3) - 1\right]\boldsymbol{X}\boldsymbol{X}' \tag{9.4.39}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta}_3 \partial \sigma_3} = \sum_3 \left[C(B_3)\frac{\varrho_3^2}{\sigma_3^3}\frac{u_3}{1-\varrho_3^2} + \frac{\varrho_3}{\sigma_3^2}\frac{\lambda(B_3)}{\sqrt{1-\varrho_3^2}} - 2\frac{u_3}{\sigma_3^3}\right]\boldsymbol{X} \tag{9.4.40}$$

$$\frac{\partial^2 l}{\partial \boldsymbol{\beta}_3 \partial \varrho_3} = \sum_3 \left[-C(B_3)\frac{\frac{u_3}{\sigma_3} + \varrho_3 \boldsymbol{Z}'\boldsymbol{\gamma}}{(1-\varrho_3^2)^2}\frac{\varrho_3}{\sigma_3} - \frac{\lambda(B_3)}{\sigma_3}\frac{1}{(1-\varrho_3^2)^{\frac{3}{2}}}\right]\boldsymbol{X} \tag{9.4.41}$$

$$\begin{aligned}\frac{\partial^2 l}{\partial \sigma_3^2} &= \sum_3 \left[\frac{1}{\sigma_3^2} - 3\frac{u_3^2}{\sigma_3^4} + 2\lambda(B_3)\frac{y_3 - \boldsymbol{X}'\boldsymbol{\beta_3}}{\sqrt{1-\varrho_3^2}}\frac{\varrho_3}{\sigma_3^3}\right] + \\ &\quad + \sum_3 \frac{\varrho_3^2}{\sigma_3^4}\frac{u_3^2}{1-\varrho_3^2}C(B_3)\end{aligned} \tag{9.4.42}$$

$$\frac{\partial^2 l}{\partial \sigma_3 \partial \varrho_3} = -\frac{1}{(1-\varrho_3^2)^{\frac{3}{2}}}\sum_3 \frac{u_3}{\sigma_3^2}\left[C(B_3)\frac{\varrho_3\left(\frac{u_3}{\sigma_3} + \varrho_3 \boldsymbol{Z}'\boldsymbol{\gamma}\right)}{\sqrt{1-\varrho_3^2}} + \lambda(B_3)\right] \tag{9.4.43}$$

$$\begin{aligned}\frac{\partial^2 l}{\partial \varrho_3^2} &= \sum_3 C(B_3)\left[\frac{\frac{1}{\sigma_3}u_3 + \varrho_3 \boldsymbol{Z}'\boldsymbol{\gamma}}{(1-\varrho_3^2)^{\frac{3}{2}}}\right]^2 + \\ &\quad + \sum_3 \lambda(B_3)\frac{\boldsymbol{Z}'\boldsymbol{\gamma}(1+2\varrho_3^2) + 3\varrho_3\frac{1}{\sigma_3}u_3}{(1-\varrho_3^2)^{\frac{7}{2}}}\end{aligned} \tag{9.4.44}$$

Siin on tähistatud

$$B_2 = -\frac{\boldsymbol{Z}'\boldsymbol{\gamma} + \frac{\varrho_2}{\sigma_2}(y_2 - \boldsymbol{X}'\boldsymbol{\beta_2})}{\sqrt{1-\varrho_2^2}} \tag{9.4.45}$$

$$B_3 = \frac{\boldsymbol{Z}'\boldsymbol{\gamma} + \frac{\varrho_3}{\sigma_3}(y_3 - \boldsymbol{X}'\boldsymbol{\beta_3})}{\sqrt{1-\varrho_3^2}} \tag{9.4.46}$$

$$\lambda(B) = \frac{\phi(B)}{\Phi(B)} \tag{9.4.47}$$

$$u_2 = y_2 - \boldsymbol{X}'\boldsymbol{\beta_2} \tag{9.4.48}$$

$$u_3 = y_3 - \boldsymbol{X}'\boldsymbol{\beta_3} \tag{9.4.49}$$

$$C(B) = -\frac{\Phi(B)\phi(B)B + \phi(B)^2}{\Phi(B)^2} \tag{9.4.50}$$

9.5 Kestusmudelid

Tähistused:

τ kestus, algseisundis viibitud aeg

t kalendriaeg

9.5.1 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

$$\hat{h} = \frac{n_j}{r_j} \tag{9.5.1}$$

$$\widehat{\mathrm{Var}\,\hat{h}_j} = \frac{\hat{h}_j(1-\hat{h}_j)}{r_j}. \tag{9.5.2}$$

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

$$\hat{\vartheta}_j = 1 - (1-\hat{h}_j)^{1/k} \tag{9.5.3}$$

$$\widehat{\mathrm{Var}\,\hat{\vartheta}_j} = \frac{\mathrm{Var}\,\hat{h}_j}{\left[k(1-\hat{h}_j)^{1-1/k}\right]^2} \tag{9.5.4}$$

KM hinnang pidevas ajas Lahkugu aja t jooksul r algseisundis olnud inimesest n. Keskmine hasart ajaühikus on

$$\hat{\vartheta} = -\frac{1}{t}\log(1-n/r) \tag{9.5.5}$$

$$\widehat{\mathrm{Var}\,\hat{\vartheta}} = \frac{1}{t^2}\frac{n/r}{r-n}. \tag{9.5.6}$$

9.5.2 Multiplikatiivne mittevaadeldav heterogeensus

Eeldame et hasart avaldub

$$\vartheta(\tau|x,v) = \lambda(\tau|x)v \tag{9.5.7}$$

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

$$\mathbb{E}(v|T \geq \tau) = -\frac{\mathcal{L}'[z(\tau|x)]}{\mathcal{L}[z(\tau|x)]} \tag{9.5.8}$$

kus on integreeritud hasart v-d arvestamata:

$$z(\tau|x) = \int_0^\tau \lambda(s|x)\,\mathrm{d}s \tag{9.5.9}$$

Kui v on algseisundisse sissevoolus ühikdispersiooniga gammajaotus parameetriga α, siis

$$\mathbb{E}(v|T \geq \tau) = \frac{\alpha}{z(\tau|x) + \alpha^{1/2}}. \tag{9.5.10}$$

9.5.3 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 $\boldsymbol{\lambda}$, kusjuures ajavahemiku j põhihasart olgu e^{λ_j}. Mittevaadeldav heterogeensus on diskreetse jaotusega:

$$v = \begin{cases} v_h \equiv 1, & \text{tõenäosusega} \quad p_h, \\ v_l, & \text{tõenäosusega} \quad p_l = 1 - p_h. \end{cases} \tag{9.5.11}$$

Sobiv on parametriseerida

$$\begin{array}{lcl} v_1 & = & \mathrm{e}^{\tilde{v}_1} \\ \dots & & \\ v_K & = & 1 \end{array} \qquad \text{ja} \qquad \begin{array}{lcl} p_1 & = & \Lambda(\tilde{p}_1) \\ \dots & & \\ p_K & = & 1 - \sum^{K-1} p_k, \end{array} \tag{9.5.12}$$

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 $\mu_i = \mathrm{e}^{\gamma x_i'}$ olgu hasardi inimesest sõltuv osa. T_{ij} olgu teadaolev (võimalik et tsenseeritud) aeg, mis inimene i veetis uuritavas seisundis ajaperioodi j jooksul. Vektor $\boldsymbol{T}_i$ on M-vektor, mille komponendid on T_{ij} ning vector $\boldsymbol{d}_i$ 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:

$$\mathcal{L}_i \quad = \quad p_l \mathcal{L}_{li} + p_h \mathcal{L}_{hi} \quad = \quad p_l \left(v_l \mu_i \mathrm{e}^{\lambda_{m_i}}\right)^{\delta_i} \mathrm{e}^{-z_{li}} + p_h \left(\mu_i \mathrm{e}^{\lambda_{m_i}}\right)^{\delta_i} \mathrm{e}^{-z_{hi}}, \tag{9.5.13}$$

kus

$$z_{li} = v_l \mu_i \sum_{j=1}^{M-1} \mathrm{e}^{\lambda_j} T_{ij} \qquad \text{ja} \qquad z_{hi} = \mu_i \sum_{j=1}^{M-1} \mathrm{e}^{\lambda_j} T_{ij} \tag{9.5.14}$$

on integreeritud hasart. Log-laiklihuudi gradient avaldub:

$$\frac{\partial \ell_i}{\partial v_l} = \frac{p_l \mathcal{L}_{li}}{\mathcal{L}} \left[\frac{\delta_i}{v_l} - z_{hi}\right] \tag{9.5.15}$$

$$\frac{\partial \ell_i}{\partial p_l} = \frac{1}{\mathcal{L}_i} \left[\mathcal{L}_{li} - \mathcal{L}_{hi}\right] \tag{9.5.16}$$

$$\frac{\partial \ell_i}{\partial \boldsymbol{\lambda}} = p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i} \left(\delta_i \boldsymbol{d}_i - v_l \mu_i \boldsymbol{T}_i\right) + p_h \frac{\mathcal{L}_{hi}}{\mathcal{L}_i} \left(\delta_i \boldsymbol{d}_i - \mu_i \boldsymbol{T}_i\right) \tag{9.5.17}$$

$$\frac{\partial \ell_i}{\partial \boldsymbol{\gamma}} = \frac{\boldsymbol{x_i}}{\mathcal{L}_i} \left[p_l \mathcal{L}_l \left(\delta_i - z_{li}\right) + p_h \mathcal{L}_h \left(\delta_i - z_{hi}\right)\right] \tag{9.5.18}$$

$$\tag{9.5.19}$$

ja hessi maatriks:

$$\frac{\partial^2 \ell_i}{\partial v_l^2} = \frac{\mathcal{L}_{li}}{\mathcal{L}_i} \left[\left(\frac{\delta_i}{v_l} - z_{hi}\right)^2 \left(1 - p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i}\right) - \frac{\delta_i}{v_l^2}\right] \tag{9.5.20}$$

$$\frac{\partial^2 \ell_i}{\partial p_l^2} = -\left(\frac{\mathcal{L}_{li}}{\mathcal{L}_i} - \frac{\mathcal{L}_{hi}}{\mathcal{L}_i}\right)^2 \tag{9.5.21}$$

$$\frac{\partial^2 \ell_i}{\partial v_l \partial p_l} = \frac{\mathcal{L}_{li}}{\mathcal{L}_i}\left(\frac{\delta_i}{v_l} - z\right)\left(1 - p_l\frac{\mathcal{L}_{li} - \mathcal{L}_{hi}}{\mathcal{L}_i}\right) \tag{9.5.22}$$

$$\begin{aligned}
\frac{\partial^2 \ell_i}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} &= \frac{p_l}{\mathcal{L}_i}\left[\frac{\partial \mathcal{L}_{li}}{\partial \lambda}(\delta_i \boldsymbol{d}_i - v_l \mu_i \boldsymbol{T}_i) - \mathcal{L}_{li}\mathrm{diag}\left(v_l \mu_i \mathrm{e}^{\boldsymbol{\lambda}} * \boldsymbol{T}_i\right)\right] + \\
&+ \frac{p_h}{\mathcal{L}_i}\left[\frac{\partial \mathcal{L}_{hi}}{\partial \lambda}(\delta_i \boldsymbol{d}_i - \mu_i \boldsymbol{T}_i) - \mathcal{L}_{hi}\mathrm{diag}\left(\mu_i \mathrm{e}^{\boldsymbol{\lambda}} * \boldsymbol{T}_i\right)\right] - \\
&- \frac{1}{\mathcal{L}_i^2}\left(\frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\lambda}}\right)^2
\end{aligned} \tag{9.5.23}$$

$$\begin{aligned}
\frac{\partial^2 \ell_i}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} &= p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i}\left[(\delta_i - z_{li})^2 - z_{li}\right]\boldsymbol{x}_i\boldsymbol{x}_i' + p_h \frac{\mathcal{L}_{hi}}{\mathcal{L}_i}\left[(\delta_i - z_{hi})^2 - z_{hi}\right]\boldsymbol{x}_i\boldsymbol{x}_i' - \\
&- \frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\gamma}}\frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\gamma}'}
\end{aligned} \tag{9.5.24}$$

$$\begin{aligned}
\frac{\partial^2 \ell_i}{\partial \boldsymbol{\lambda} \boldsymbol{\gamma}'} &= \left[(\delta_i - z_{li})(\delta_i \boldsymbol{d}_i - z_{li}) - z_{li}\right]\frac{\mathcal{L}_{li}}{\mathcal{L}_i}\boldsymbol{x}_i' p_l + \\
&+ \left[(\delta_i - z_{hi})(\delta_i \boldsymbol{d}_i - z_{hi}) - z_{hi}\right]\frac{\mathcal{L}_{hi}}{\mathcal{L}_i}\boldsymbol{x}_i' p_h - \frac{1}{\mathcal{L}_i}\frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\lambda}}\frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\gamma}'}
\end{aligned} \tag{9.5.25}$$

$$\begin{aligned}
\frac{\partial^2 \ell_i}{\partial \boldsymbol{\lambda} \partial v_l} &= p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i}\left(\frac{\delta_i}{v_l} - z_{hi}\right)\left(\delta_i \boldsymbol{d}_i - v_l \mu_i \mathrm{e}^{\boldsymbol{\lambda}} * \boldsymbol{T}_i - \frac{1}{\mathcal{L}_i}\frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\lambda}}\right) - \\
&- p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i}\mu_i \mathrm{e}^{\boldsymbol{\lambda}} * \boldsymbol{T}_i
\end{aligned} \tag{9.5.26}$$

$$\begin{aligned}
\frac{\partial^2 \ell_i}{\partial \boldsymbol{\lambda} \partial p_l} &= \frac{\mathcal{L}_{li}}{\mathcal{L}_i}\left(\delta_i d_i - v_l \mu_i \mathrm{e}^{\boldsymbol{\lambda}} * \boldsymbol{T}_i\right) - \frac{\mathcal{L}_{hi}}{\mathcal{L}_i}\left(\delta_i d_i - \mu_i \mathrm{e}^{\boldsymbol{\lambda}} * \boldsymbol{T}_i\right) - \\
&- \frac{\mathcal{L}_{li} - \mathcal{L}_{hi}}{\mathcal{L}_i}\frac{1}{\mathcal{L}_i}\frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\lambda}}
\end{aligned} \tag{9.5.27}$$

$$\begin{aligned}
\frac{\partial^2 \ell_i}{\partial \boldsymbol{\gamma} \partial v_l} &= p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i}\left(\frac{\delta_i}{v_l} - z_{hi}\right)(\delta_i - z_{li})\boldsymbol{x}_i - \\
&- p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i}\left(\frac{\delta_i}{v_l} - z_{hi}\right)\frac{1}{\mathcal{L}_i}\frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\gamma}}\boldsymbol{x}_i - p_l \frac{\mathcal{L}_{li}}{\mathcal{L}_i} z_i \boldsymbol{x}_i
\end{aligned} \tag{9.5.28}$$

$$\frac{\partial^2 \ell_i}{\partial \boldsymbol{\gamma} \partial p_l} = \left[\frac{\mathcal{L}_{li}}{\mathcal{L}_i}(\delta_i - z_{li}) - \frac{\mathcal{L}_{hi}}{\mathcal{L}_i}(\delta_i - z_{hi})\right]\boldsymbol{x}_i - \frac{\mathcal{L}_{li} - \mathcal{L}_{hi}}{\mathcal{L}_i}\frac{1}{\mathcal{L}_i}\frac{\partial \mathcal{L}_i}{\partial \boldsymbol{\gamma}} \tag{9.5.29}$$

Eelnevas tähendab $\{\mathrm{e}^{\boldsymbol{\lambda}}\}_i = \mathrm{e}^{\lambda_i}$, $*$ vektorite elementide kaupa korrutamist ($\{\boldsymbol{a} * \boldsymbol{b}\}_i = a_i b_i$) ning $\mathrm{diag}\,\boldsymbol{a}$ on maatriks, mille peadiagonaalil on vektor $\boldsymbol{a}$ 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

$$\theta(t|\boldsymbol{x}, v) = v\mathrm{e}^{\lambda(t)}\mathrm{e}^{\gamma\boldsymbol{x}_i'}. \tag{9.5.30}$$

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

$$\mathcal{L}_{ij}(\cdot|v) = v\theta(t_{ij}|\boldsymbol{x}_{ij})^{\delta_{ij}} S(t_{ij}|\boldsymbol{x}_{ij}, v). \tag{9.5.31}$$

Siin t_{ij} on spelli vaadeldud kestus, δ_{ij} on mitte-tsenseerituse indikaator ja $\boldsymbol{x}_{ij}$ 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

$$\mathcal{L}_i(\cdot|v) = \prod_{j=1}^{N_i} \mathcal{L}_{ij}(\cdot|v). \tag{9.5.32}$$

Vaadeldav laiklihuud avaldub:

$$\mathcal{L}_i(\cdot) = \sum_{k=1}^{K} p_k \mathcal{L}_i(\cdot|v_k). \tag{9.5.33}$$

Log-laiklihuudi gradient avaldub:

$$\frac{\partial \ell_i}{\partial v_k} = \frac{p_k}{\mathcal{L}_i(\cdot)} \sum_j \frac{\mathcal{L}_i(\cdot|v_k)}{\mathcal{L}_{ij}(\cdot|v_k)} \frac{\partial \mathcal{L}_{ij}(\cdot|v_k)}{\partial v_k} \tag{9.5.34}$$

$$\frac{\partial \ell_i}{\partial p_l} = \frac{\mathcal{L}_i(\cdot|v_k)}{\mathcal{L}_i(\cdot)} \tag{9.5.35}$$

$$\frac{\partial \ell_i}{\partial \boldsymbol{\lambda}} = \frac{1}{\mathcal{L}_i(\cdot)} \sum_k p_k \sum_j \frac{\mathcal{L}_i(\cdot|v_k)}{\mathcal{L}_{ij}(\cdot|v_k)} \frac{\partial \mathcal{L}_{ij}(\cdot|v_k)}{\partial \boldsymbol{\lambda}} \tag{9.5.36}$$

$$\frac{\partial \ell_i}{\partial \boldsymbol{\gamma}} = \frac{1}{\mathcal{L}_i(\cdot)} \sum_k p_k \sum_j \frac{\mathcal{L}_i(\cdot|v_k)}{\mathcal{L}_{ij}(\cdot|v_k)} \frac{\partial \mathcal{L}_{ij}(\cdot|v_k)}{\partial \boldsymbol{\gamma}}. \tag{9.5.37}$$

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

9.5.4 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: $[0, t_1), [t_1, t_2), \ldots, [t_{T-1}, t_T), [t_T, \infty)$. Iga indiviidi i kohta olgu vaadeldav et millises vahemikus ta algsest seisundist lahkus, või et millises vahemikust vaatlus on tsenseeritud. Eeldame MPH mudelit nagu 9.5.3. osas:

$$\vartheta(\tau|\boldsymbol{x}, v) = \lambda(\tau)\mathrm{e}^{\boldsymbol{\beta}'\boldsymbol{x}}v. \tag{9.5.38}$$

Tõenäosus, et isik jääb kogu intervalli n jooksul algseisundisse avaldub

$$S_n(\boldsymbol{x}, v) = \exp(-vz_n(\boldsymbol{x})) = \exp(-v\mathrm{e}^{\boldsymbol{\beta}'\boldsymbol{x}} \int_{\tau_{n-1}}^{\tau_n} \lambda(s)\,\mathrm{d}s). \tag{9.5.39}$$

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

$$\mathrm{e}^{\tilde{\lambda}_n}(t_n - t_{n-1}) \equiv \int_{\tau_{n-1}}^{\tau_n} \lambda(s)\,\mathrm{d}s, \tag{9.5.40}$$

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

$$\mathcal{L}(n|\boldsymbol{x}, v) = \left(1 - \mathrm{e}^{-vz_n(\boldsymbol{x})}\right)^{\delta} \prod_{m=1}^{n-1} \mathrm{e}^{-vz_m(\boldsymbol{x})}, \tag{9.5.41}$$

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

$$\ell(n|\boldsymbol{x}) = \log\left(\sum_{k=1}^{K} p_k \mathcal{L}(n|\boldsymbol{x}, v_k)\right). \tag{9.5.42}$$

v-spetsiifilise laiklihuudi gradient avaldub

$$\frac{\partial}{\partial\boldsymbol{\beta}}\mathcal{L}(n|\boldsymbol{x}, v) = vE_g(n|\boldsymbol{x}, v)\frac{\partial}{\partial\boldsymbol{\beta}}z_n(\boldsymbol{x}) - vS_g(n|\boldsymbol{x}, v)\sum_{m=1}^{n-1}\frac{\partial}{\partial\boldsymbol{\beta}}z_m(\boldsymbol{x}) \tag{9.5.43}$$

$$\frac{\partial}{\partial\tilde{\lambda}_s}\mathcal{L}(n|\boldsymbol{x}, v) = vE_g(n|\boldsymbol{x}, v)\frac{\partial}{\partial\tilde{\lambda}_s}z_n(\boldsymbol{x}) - vS_g(n|\boldsymbol{x}, v)\sum_{m=1}^{n-1}\frac{\partial}{\partial\tilde{\lambda}_s}z_m(\boldsymbol{x}) \tag{9.5.44}$$

$$\frac{\partial}{\partial v}\mathcal{L}(n|\boldsymbol{x}, v) = E_g(n|\boldsymbol{x}, v)z_n(\boldsymbol{x}) - S_g(n|\boldsymbol{x}, v)\sum_{m=1}^{n-1}z_m(\boldsymbol{x}) \tag{9.5.45}$$

kus

$$E_g(n|\boldsymbol{x}, v) = \delta\mathrm{e}^{-vz_n(\boldsymbol{x})}\prod_{m=1}^{n-1}\mathrm{e}^{-vz_m(\boldsymbol{x})} \tag{9.5.46}$$

$$S_g(n|\boldsymbol{x}, v) = \mathcal{L}(n|\boldsymbol{x}, v) = \left(1 - \mathrm{e}^{-vz_n(\boldsymbol{x})}\right)^{\delta} \prod_{m=1}^{n-1} \mathrm{e}^{-vz_m(\boldsymbol{x})} \tag{9.5.47}$$

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

$$\frac{\partial}{\partial \boldsymbol{\beta}} z_n(\boldsymbol{x}) = z_n(\boldsymbol{x})\boldsymbol{x}_n \tag{9.5.48}$$

$$\frac{\partial}{\partial \tilde{\lambda}_s} z_n(\boldsymbol{x}) = z_n(\boldsymbol{x})\mathbb{1}(s = n). \tag{9.5.49}$$

Kogulaiklihuudi gradient on

$$\frac{\partial}{\partial \boldsymbol{\beta}} \ell(n|\boldsymbol{x}) = \frac{1}{\mathcal{L}(n|\boldsymbol{x})} \left(\sum_{k=1}^{K} p_k \frac{\partial}{\partial \boldsymbol{\beta}} \mathcal{L}(n|\boldsymbol{x}, v) \right) \tag{9.5.50}$$

$$\frac{\partial}{\partial \tilde{\lambda}_s} \ell(n|\boldsymbol{x}) = \frac{1}{\mathcal{L}(n|\boldsymbol{x})} \left(\sum_{k=1}^{K} p_k \frac{\partial}{\partial \tilde{\lambda}_s} \mathcal{L}(n|\boldsymbol{x}, v) \right) \tag{9.5.51}$$

$$\frac{\partial}{\partial v_k} \ell(n|\boldsymbol{x}) = \frac{1}{\mathcal{L}(n|\boldsymbol{x})} p_k \frac{\partial}{\partial v_k} \mathcal{L}(n|\boldsymbol{x}, v) \tag{9.5.52}$$

$$\frac{\partial}{\partial p_k} \ell(n|\boldsymbol{x}) = \frac{\mathcal{L}(n|\boldsymbol{x}, v_k)}{\mathcal{L}(n|\boldsymbol{x})} \tag{9.5.53}$$

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

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

$$\mathcal{L}_{ij}^m(\cdot|v^m) = \left[\vartheta^m(\tau_{ij}|\boldsymbol{x}_{ij})\right]^{\delta_{ij}^m} S^m(\tau_{ij}|\boldsymbol{x}_{ij}, v^m). \tag{9.5.54}$$

Siirdega seotud liige $\left[\vartheta^m(\tau_{ij}|\boldsymbol{x}_{ij})\right]^{\delta_{ij}^m}$ ja püsimisega seotud liige $S^m(\tau_{ij}|\boldsymbol{x}_{ij}, v^m)$ avalduvad nii nagu ühe spelli ja ühe lõppseisundi puhul (vaata näiteks 9.5.3 või 9.5.4). Siin τ_{ij} on spelli vaadeldud kestus, δ_{ij}^m on mitte-tsenseerituse indikaator m-seisundi mõttes (= 1 kui läks seisundisse m ja 0 kui es lähe) ja $\boldsymbol{x}_{ij}$ on inimese i vaadeldavad isikutunnused spelli j ajal. Spelli kogulaiklihuud on:

$$\mathcal{L}_{ij}(\cdot|\boldsymbol{x}_{ij}, \boldsymbol{v}) = \prod_{m=1}^{M} \mathcal{L}_{ij}^m(\cdot|\boldsymbol{x}_{ij}, v^m). \tag{9.5.55}$$

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

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

$$\mathcal{L}_i(\cdot|\boldsymbol{x}, \boldsymbol{v}) = \prod_{j=1}^{N_i} \mathcal{L}_{ij}(\cdot|\boldsymbol{x}, \boldsymbol{v}). \tag{9.5.56}$$

ja vaadeldav laiklihuud avaldub:

$$\mathcal{L}_i(\cdot|\boldsymbol{x}) = \sum_{k^1=1}^{K^1} \cdots \sum_{k^M=1}^{K^M} p_{k^1 \ldots k^M} \mathcal{L}_i(\cdot|\boldsymbol{x}, \boldsymbol{v}). \tag{9.5.57}$$

Laiklihuudi gradient avaldub:

$$\frac{\partial \mathcal{L}_i(\cdot|\boldsymbol{x})}{\partial \boldsymbol{\lambda}^m} = \sum_{k^1=1}^{K^1} \cdots \sum_{k^M=1}^{K^M} p_{k^1 \ldots k^M} \mathcal{L}_i(\cdot|\boldsymbol{x}, \boldsymbol{v}) \sum_{j=1}^{N_i} \left[\frac{1}{\mathcal{L}_{ij}^m(\cdot|\boldsymbol{x} v_{k^m}^m)} \frac{\partial \mathcal{L}_{ij}^m(\cdot|\boldsymbol{x} v_{k^m}^m)}{\partial \boldsymbol{\lambda}^m} \right] \tag{9.5.58}$$

$$\frac{\partial \mathcal{L}_i(\cdot|\boldsymbol{x})}{\partial \boldsymbol{\gamma}^m} = \sum_{k^1=1}^{K^1} \cdots \sum_{k^M=1}^{K^M} p_{k^1 \ldots k^M} \mathcal{L}_i(\cdot|\boldsymbol{x}, \boldsymbol{v}) \sum_{j=1}^{N_i} \left[\frac{1}{\mathcal{L}_{ij}^m(\cdot|\boldsymbol{x} v_{k^m}^m)} \frac{\partial \mathcal{L}_{ij}^m(\cdot|\boldsymbol{x} v_{k^m}^m)}{\partial \boldsymbol{\gamma}^m} \right] \tag{9.5.59}$$

$$\frac{\partial \mathcal{L}_i(\cdot|\boldsymbol{x})}{\partial v_k^m} = \sum_{k^1=1}^{K^1} \cdots \sum_{k^{m-1}=1}^{K^{m-1}} \sum_{k^{m+1}=1}^{K^{m+1}} \cdots \sum_{k^M=1}^{K^M} p_{k^1 \ldots k^{m-1} k k^{m+1} \ldots k^M} \cdot$$

$$\cdot \mathcal{L}_i(\cdot|\boldsymbol{x}, \boldsymbol{v}) \sum_{j=1}^{N_i} \left[\frac{1}{\mathcal{L}_{ij}^m(\cdot|\boldsymbol{x} v_k^m)} \frac{\partial \mathcal{L}_{ij}^m(\cdot|\boldsymbol{x} v_k^m)}{\partial \gamma^m} \right] \tag{9.5.60}$$

$$\frac{\partial \mathcal{L}_i(\cdot|\boldsymbol{x})}{\partial p_{k^1 \dots k^M}} = \mathcal{L}_i(\cdot|\boldsymbol{x}, (v_{k^1}^1, \dots, v_{k^M}^M)) \tag{9.5.61}$$

$\mathcal{L}_{ij}^m(\cdot|\boldsymbol{x}, \boldsymbol{v})$ tuletised on nii nagu sõltumatute vaatluste korral.

9.5.5 Parametriseerimine

Logistic transition probability in discrete-time In discrete time, it is useful to parameterise the destination-specific exit probabilities logistically as

$$p^m \equiv \Pr(\text{exit to destination } m) = \frac{e^{\lambda^m}}{1 + \sum_k e^{\lambda^k}}. \tag{9.5.62}$$

The probabilities are guaranteed to be positive and their sum to be less than one while λ-s may be unbounded. The components of the corresponding gradient transformation matrix are

$$\frac{\partial}{\partial \lambda^k} p^m = \mathbb{1}(k = m) p_m - p_k p_m \tag{9.5.63}$$

and the inverse transformation

$$l^m = \log \frac{p^m}{1 - \sum_k p^k}. \tag{9.5.64}$$

Transition probability in continuous time A good choice for parametrising the time-dependent part of the hazard is

$$\lambda = e^{\tilde{\lambda}} \tag{9.5.65}$$

The λ is now guaranteed to be positive.

Diskreetne mittevaadeldav heterogeensus Diskreetne mittevaadeldav heterogeensus, üks lõppseisund: $v \in \{v_1, v_2, \dots, v_K$ ja vastavad tõenäousused $p_1, p_2, \dots, p_K$.

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:

$$\begin{pmatrix} \frac{\partial}{\partial \tilde{\boldsymbol{v}}} \ell_i \\ \frac{\partial}{\partial \tilde{\boldsymbol{p}}} \ell_i \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial \tilde{\boldsymbol{v}}} v & \frac{\partial}{\partial \tilde{\boldsymbol{v}}} p \\ \frac{\partial}{\partial \tilde{\boldsymbol{p}}} v & \frac{\partial}{\partial \tilde{\boldsymbol{p}}} p \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial v} \ell_i \\ \frac{\partial}{\partial p} \ell_i \end{pmatrix} \equiv \mathsf{C} \begin{pmatrix} \frac{\partial}{\partial v} \ell_i \\ \frac{\partial}{\partial p} \ell_i \end{pmatrix} \tag{9.5.66}$$

Maatriksi 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:

$$\mathsf{\Sigma} = \mathsf{C}' \tilde{\mathsf{\Sigma}} \mathsf{C} \tag{9.5.67}$$

Heterogeensus Tõenäosused on võimalik parametriseerida kui

$$p_k = \frac{e^{\tilde{p}_k}}{\sum_{k=1}^{K} e^{\tilde{p}_k}} \qquad \text{ja} \qquad \sum_{k=1}^{K} \tilde{p}_k = 0, \tag{9.5.68}$$

kus K on jaotuse toetuspunktide arv.

Keskväärtuse normeerimine v keskväärtuse võib normeerida üheks:

$$v_k = \mathrm{e}^{\tilde{v}_k} \qquad \text{ja} \qquad \sum_{k=1}^{K} p_k v_k = 1. \tag{9.5.69}$$

Vastav pöördteisendus tõenäosuste tarvis on

$$\tilde{p}_k = \log p_k - \frac{\sum_{l=1}^{K} \log p_l}{K}. \tag{9.5.70}$$

C komponendid on:

$$\frac{\partial}{\partial \tilde{v}_l} v_k = \begin{cases} v_k & \text{kui} \qquad k = l < K \\ -v_l \dfrac{p_l}{p_K} & \text{kui} \qquad k = K \\ 0 & \text{muudel juhtudel} \end{cases} \tag{9.5.71}$$

$$\frac{\partial}{\partial \tilde{v}_l} p_k = 0 \tag{9.5.72}$$

$$\frac{\partial}{\partial \tilde{p}_l} v_k = \begin{cases} \dfrac{1}{p_K} \left[(1 - p_K) + p_K v_K + (1 - p_l) v_l\right] & \text{kui} \qquad k = K \\ 0 & \text{muudel juhtudel} \end{cases} \tag{9.5.73}$$

$$\frac{\partial p_k}{\partial \tilde{p}_l} = -p_k(p_l - p_K) + \mathbb{1}(k = l) p_K - \mathbb{1}(K = k) p_K \tag{9.5.74}$$

v **komponendi normeerimine** Defineerime

$$v_K = 1. \tag{9.5.75}$$

Näib et nii on intervallandmete juures tulemused mõnevõrra stabiilsemad, samas on põhihasarti raskem tõlgendada.

C komponendid on

$$\frac{\partial}{\partial \tilde{v}_l} v_k = \begin{cases} v_k & \text{kui} \qquad k = l < K \\ 0 & \text{muudel juhtudel} \end{cases} \tag{9.5.76}$$

$$\frac{\partial}{\partial \tilde{v}_l} p_k = 0 \tag{9.5.77}$$

$$\frac{\partial}{\partial \tilde{p}_l} v_k = 0 \tag{9.5.78}$$

$$\frac{\partial p_k}{\partial \tilde{p}_l} = -\frac{\mathrm{e}^{\tilde{p}_k} \left(\mathrm{e}^{\tilde{p}_l} - \mathrm{e}^{\tilde{p}_H}\right)}{\left(\sum_{i=1}^{H} \mathrm{e}^{\tilde{p}_i}\right)^2} + \mathbb{1}(k = l) \frac{\mathrm{e}^{\tilde{p}_k}}{\sum_{i=1}^{H} \mathrm{e}^{\tilde{p}_i}} + \mathbb{1}(K = k) \frac{\mathrm{e}^{\tilde{p}_K}}{\sum_{i=1}^{K} \mathrm{e}^{\tilde{p}_i}} \tag{9.5.79}$$

M **lõppseisundit** (9.5.66) asemel võib nüüd kirjutada:

$$\begin{pmatrix} \frac{\partial}{\partial \tilde{\boldsymbol{v}}^1}\ell_i \\ \frac{\partial}{\partial \tilde{\boldsymbol{v}}^2}\ell_i \\ \cdots \\ \frac{\partial}{\partial \tilde{\boldsymbol{v}}^M}\ell_i \\ \frac{\partial}{\partial \tilde{\boldsymbol{p}}}\ell_i \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial \tilde{\boldsymbol{v}}^1}v^1 & \frac{\partial}{\partial \tilde{\boldsymbol{v}}^1}v^2 & \cdots & \frac{\partial}{\partial \tilde{\boldsymbol{v}}^1}v^M & \frac{\partial}{\partial \tilde{\boldsymbol{v}}^1}p \\ \frac{\partial}{\partial \tilde{\boldsymbol{v}}^2}v^1 & \frac{\partial}{\partial \tilde{\boldsymbol{v}}^2}v^2 & \cdots & \frac{\partial}{\partial \tilde{\boldsymbol{v}}^2}v^M & \frac{\partial}{\partial \tilde{\boldsymbol{v}}^2}p \\ \cdots & \cdots & \ddots & \cdots & \cdots \\ \frac{\partial}{\partial \tilde{\boldsymbol{v}}^M}v^1 & \frac{\partial}{\partial \tilde{\boldsymbol{v}}^M}v^2 & \cdots & \frac{\partial}{\partial \tilde{\boldsymbol{v}}^M}v^M & \frac{\partial}{\partial \tilde{\boldsymbol{v}}^M}p \\ \frac{\partial}{\partial \tilde{\boldsymbol{p}}}v^1 & \frac{\partial}{\partial \tilde{\boldsymbol{p}}}v^2 & \cdots & \frac{\partial}{\partial \tilde{\boldsymbol{p}}}v^M & \frac{\partial}{\partial \tilde{\boldsymbol{p}}}p \end{pmatrix} \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{v}^1}\ell_i \\ \frac{\partial}{\partial \boldsymbol{v}^2}\ell_i \\ \cdots \\ \frac{\partial}{\partial \boldsymbol{v}^M}\ell_i \\ \frac{\partial}{\partial \boldsymbol{p}}\ell_i \end{pmatrix}$$

$$\equiv \mathsf{C} \begin{pmatrix} \frac{\partial}{\partial \boldsymbol{v}^1}\ell_i \\ \frac{\partial}{\partial \boldsymbol{v}^2}\ell_i \\ \cdots \\ \frac{\partial}{\partial \boldsymbol{v}^M}\ell_i \\ \frac{\partial}{\partial \boldsymbol{p}}\ell_i \end{pmatrix}, \qquad (9.5.80)$$

kus C on $\tilde{N} \times N$ maatriks.

Olgu p^m_{Dk} suuna m spetsiifiline tõenäosus, s.t. millise tõenäosusega esineb väärtus v^m_k. Tuletised võib nüüd avaldada kui

$$\frac{\partial}{\partial \tilde{p}}v^m = \frac{\partial}{\partial \tilde{p}}p\frac{\partial}{\partial p}p^m_D\frac{\partial}{\partial p^m_D}v^m \qquad (9.5.81)$$

10 Algorithms

10.1 Arrays

Array indexing Let M be a N-dimensional column-major zero-based array, and $m[i_1, i_2, \ldots, i_N]$ its element where $i_j \in \{0, \ldots, K_j - 1\}$ and K_j is the size of dimension j. Vectorized index for element $m[i_1, i_2, \ldots, i_N]$ is

$$j = i_1 + K_1 i_2 + K_2 i_3 + \cdots + K_{N-1} i_N. \tag{10.1.1}$$

The array indices can be recovered from the vector index j as:

$$\begin{aligned}
i_1 &= j \mod K_1 \\
i_2 &= [j \mod (K_1 K_2)] \,//\, K_1 \\
i_3 &= [j \mod (K_1 K_2 K_3)] \,//\, (K_1 K_2) \\
i_4 &= [j \mod (K_1 K_2 K_3 K_4)] \,//\, (K_1 K_2 K_3) \\
&\ldots
\end{aligned}$$

where $//$ is the integer division.

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.

Joe, H. and Zhu, R. (2005) Generalized poisson distribution: the property of mixture of poisson and comparison with negative binomial distribution, *Biometrical Journal*, **47**, 219–229.

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.

Tuenter, H. J. H. (2000) On the generalized poisson distribution, *Statistica Neerlandica*, **54**, 374–376.

Index