



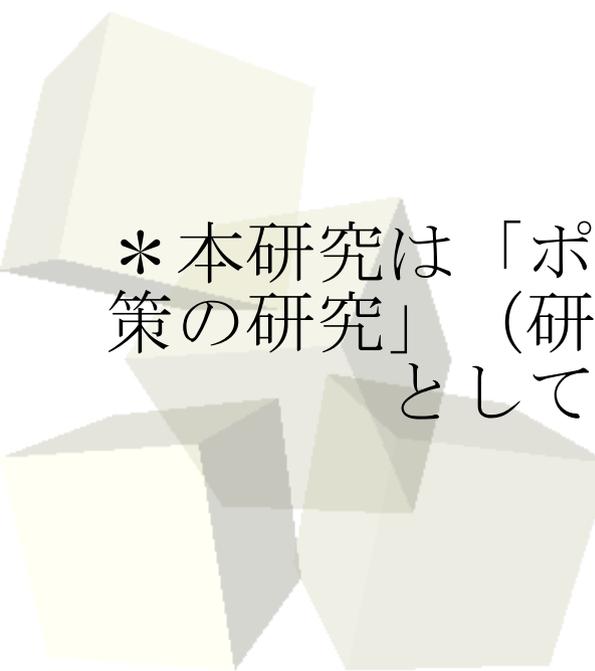
人口モデルに基づくウクライナの人口予測

神戸大学大学院保健学研究科国際保健学領域
(国際協力研究科兼任)

中澤 港

<minato-nakazawa@umin.net>

30 Nov. 2013 @ 統計数理研究所



*本研究は「ポスト人口転換期におけるオプティマルな対処方策の研究」(研究代表者:東京大学・梅崎昌裕准教授)の一部として実施した研究を発展させたものである

■ 人口予測の方法

- ・ たいていの場合、コーホート要因法かそのバリエーション
 - 女性の年齢別出生率，年齢別死亡率（と年齢別移動率）を予測
 - 年齢別女性人口に年齢別出生率を掛け，（両性の場合は出生性比により男女別に）翌年のゼロ歳人口を計算
 - 男女各歳人口がその年の年齢別死亡率に従って死亡した時に生残する人数（+純移動）を翌年の1歳上の人口とする
 - 繰り返し。が基本形
- ・ 将来の年齢別死亡率や年齢別出生率については，固定するか線形補外，あるいはシナリオとして与えるのが普通
- ・ 固定するとか線形補外には根拠が薄い，合理的なトレンド予測は困難。とくに出生率には政策の影響や突発的な変動の影響も大きいので，予測は難しい

■ 本研究での予測方法

- ・ 基本形は移動を無視したコーホート要因法
- ・ 将来の年齢別死亡率と年齢別出生率の予測に工夫
 - 現在まで複数年の年齢別死亡率と年齢別出生率に，比較的単純なパラメタライズドモデルを当てはめ
 - 個々のパラメータを arfima モデルなどで時系列解析
 - パラメータの推移トレンドを予測
 - 予測されたパラメータをモデルに代入し，年齢別死亡率と年齢別出生率の推移を予測

■ 元にするデータ

- ◆ 人口構造データ(コホート要因法は、ある時点の人口構造を元にして、出生、死亡、移動による変化を足していくので、元にする性・年齢別人口が必要):ウクライナの国勢調査は2001年にされたものが公開済み
<http://www.ukrcensus.gov.ua/eng/>
- ◆ 年齢別死亡率: Human Mortality Database(HMD)に含まれている、1961年から2005年までの男女別生命表から
- ◆ 女性の年齢別出生率: Human Fertility Database(HFD)に含まれている1959年から2006年までの女性の年齢別出生率
- ◆ HMD/HFDに登録されている国のデータなら同じ方法で分析可

■ ウクライナ人口の歴史的背景

- ◆ 1986 チェルノブイリ原発事故
- ◆ 1991 旧ソ連からの独立
- ◆ 2004 オレンジ革命を通じた欧米への接近
- ◆ 2010 親ロシア政権に復帰

■ 既存の人口予測

- ◆ US Census Bureau が年齢5歳階級で実施、発表



Human Mortality Database

- 年齢別死亡率は生命表という形で発表されている
- UCB の Dr. Wilmoth と Max-Planck の研究者がメンテしている "Human Mortality Database" (HMD) は、世界 37 の国と地域について、長年にわたる生命表を提供
- オーストラリア, オーストリア, ベラルーシ, ベルギー, ブルガリア, カナダ, チリ, チェコ, デンマーク, エストニア, フィンランド, フランス, ドイツ, ハンガリー, アイスランド, アイルランド, イスラエル, イタリア, 日本, ラトビア, リトアニア, ルクセンブルグ, オランダ, ニュージーランド, ノルウェー, ポーランド, ポルトガル, ロシア, スロバキア, スロベニア, スペイン, スウェーデン, スイス, 台湾, 英国, 米国, ウクライナ

The screenshot shows the website interface for the Human Mortality Database. On the left is a blue sidebar with a white 'HMD Main Menu' header. The menu items are categorized into Registration, Project, People, Methods, Data, Links, and General. The main content area has a title 'The Human Mortality Database' in blue. Below the title, it lists the Director (John R. Wilmoth) and Co-Director (Vladimir Shkolnikov) with their respective affiliations. A paragraph of text describes the database's purpose and history. At the bottom, a table lists 37 countries and regions.

HMD Main Menu

- Registration
 - New User
 - Change Password
 - User's Agreement
- Project
 - FAQ
 - Overview
 - History
- People
 - Acknowledgements
 - Research Teams
 - HMD Publications
- Methods
 - Brief Summary
 - Full Protocol
 - Special Methods
- Data
 - What's New
 - Explanatory Notes
 - Data Availability
 - Zipped Data Files
 - Citation Guidelines
- Links
 - Max Planck Institute
 - UC Berkeley
 - UC Berkeley Demography
 - Human Life Table Database
 - Canadian HMD
- General
 - Contact us

The Human Mortality Database

John R. Wilmoth, *Director* University of California, Berkeley

Vladimir Shkolnikov, *Co-Director* Max Planck Institute for Demographic Research

The Human Mortality Database (HMD) was created to provide detailed mortality and population data to researchers, students, journalists, policy analysts, and others interested in the history of human longevity. The project began as an outgrowth of earlier projects in the [Department of Demography at the University of California, Berkeley, USA](#), and at the [Max Planck Institute for Demographic Research in Rostock, Germany](#) (see [history](#)). It is the work of two teams of researchers in the USA and Germany (see [research teams](#)), with the help of financial backers and scientific collaborators from around the world (see [acknowledgements](#)).

We seek to provide open, international access to these data. At present the database contains detailed population and mortality data for the following 37 countries or areas:

Australia	Finland	Lithuania	Spain
Austria	France	Luxembourg	Sweden
Belarus	Germany	Netherlands	Switzerland
Belgium	Hungary	New Zealand	Taiwan
Bulgaria	Iceland	Norway	U.K.
Canada	Ireland	Poland	U.S.A.
Chile	Israel	Portugal	Ukraine
Czech Republic	Italy	Russia	
Denmark	Japan	Slovakia	
Estonia	Latvia	Slovenia	



Human Fertility Database

- 女性の年齢別出生率は、Human Fertility Database (HFD) から生データとして提供されている
 - HFD は Max-Planck 研究所とウィーン人口研究所によって維持管理されている
- <http://www.humanfertility.org/cgi-bin/main.php>

HFD Main Menu

Registration

- Login
- New User
- Change Password
- User Agreement

About the Project

- FAQ
- History
- Overview

People

- Research Teams
- Advisory Board
- Acknowledgements

Methods

- Methods Protocol

Data

- What's New
- Explanatory Notes
- Data Availability
- Data Formats
- Zipped Data Files
- Citation Guidelines

Links

- Max Planck Institute for Demographic Research



The Human Fertility Database

Joshua R. Goldstein
Director

Vladimir M. Shkolnikov
Co-Director

Tomas Sobotka
Co-Director

The Human Fertility Database (HFD) is a joint project of the [Max Planck Institute for Demographic Research](#) (MPIDR) in Rostock, Germany and the [Vienna Institute of Demography](#) (VID) in Vienna, Austria, based at MPIDR. We seek to provide free and user-friendly access to detailed, well-documented and high-quality data on period and cohort fertility and thus to facilitate research on changes and inter-country differences in fertility in the past and in the modern era.

The HFD is entirely based on one and the same type of initial data - officially registered birth counts by calendar year, mother's age (and/or cohort) and (whenever possible) biological birth order. These data, together with total female population exposure from the Human Mortality Database (www.mortality.org) and parity-specific female population exposure from selected population censuses, population registers, or large-scale surveys, are further processed using a uniform set of methods. The major HFD output includes detailed data on births, unconditional and conditional fertility rates, cohort and period fertility tables as well as selected aggregate indicators such as total fertility rates, mean ages at childbearing, and parity progression ratios.

We seek to provide open, international access to these data. At present, the database contains detailed period and cohort fertility data for the following countries:

Austria	France	Norway	Switzerland
Bulgaria	Germany	Portugal	Taiwan
Canada	Hungary	Russia	U.K.
Czech Republic	Japan	Slovakia	U.S.A.
Estonia	Lithuania	Slovenia	
Finland	Netherlands	Sweden	

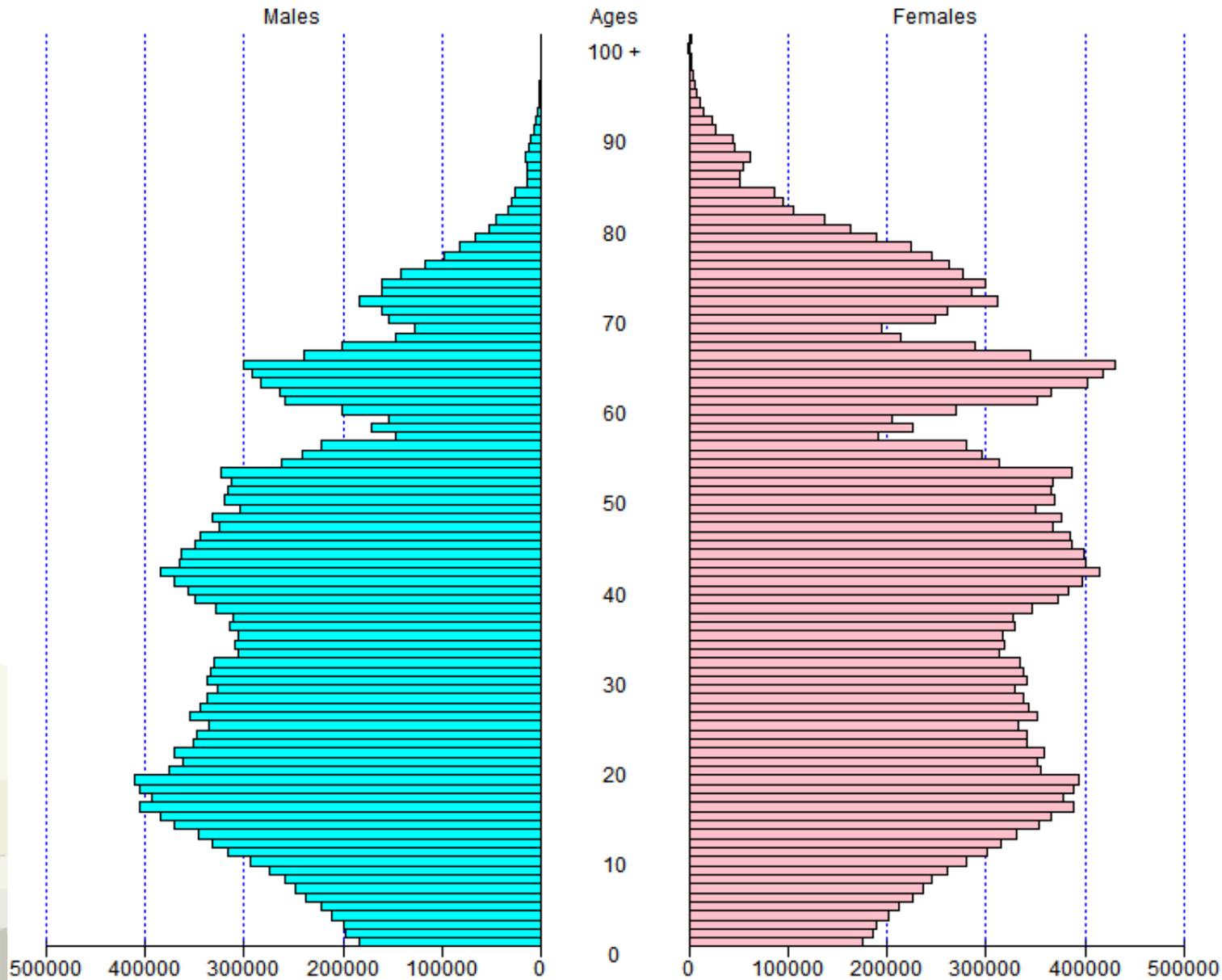
The HFD will be continually updated and more countries will be added with time. Below we present countries which are on our "coming next" list. For these countries we provide only age-specific fertility rates based on the original official data. Please be aware that these data have not been fully processed, checked, and corrected and may not be free of mistakes and biases.

Belarus	Ireland	Spain	Ukraine
---------	---------	-------	---------

- New User としてメールアドレスとパスワードを登録
- データを得る
 - ログインして欲しいデータをダウンロードし、テキスト保存して Perl などでも欲しい形式に加工するか、R で直接読む
 - HMD については、Dr. Rob J. Hyndman がメンテナをしている demography パッケージからアクセス可能。
 - `library(demography)`
 - `ukraine <- hmd.mx("UKR", "username", "passwd")`とすると、ukraine というオブジェクトにウクライナの生命表の mx が得られる。パスワードを平文でスクリプトに書いてしまうのが気になるが、便利ではある。str(ukraine) でデータ構造をみると、year, age, pop, rate からなるリスト。pop と rate はそれぞれ female, male, total からなるリスト
 - female, male, total は、それぞれ 111x51 の matrix になっている
 - HFD はまだ対応したパッケージが無い



2001年センサスによるウクライナの人口構造



```
x <- read.delim("https://sites.google.com/site/postdemotrans/progress/ukrainepop2001.txt")  
library(pyramid) # CRAN からインストール可能  
pyramid(x, Cstep=10, Laxis=0:5*100000, AxisFM="d")
```

ウクライナ生命表への死亡モデルの fitting

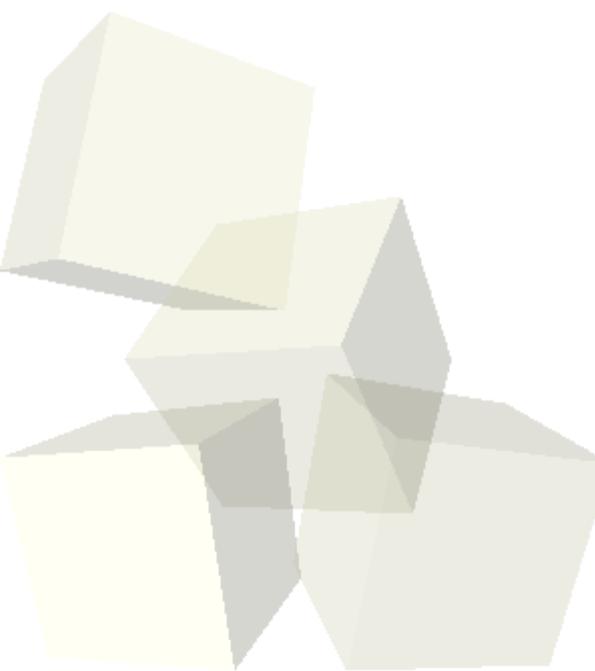
■ 生命表へのモデルの当てはめの検討課題

- ◆ どの死亡モデルを使うか？
 - Siler モデルがメカニズムを考えている中ではパラメータが 5 つと比較的少ない
 - Denny モデルが 3 パラメータで適合度がよい
- ◆ 当てはめは dx , lx , qx のどれでやるか？
- ◆ 最小化対象とする関数は、年齢別死亡について、モデルとデータの差の二乗和をとるのでよいか？
- ◆ `optim ()` の `method` は何がよいか？



モデルの fitting 方法

- Denny モデルそのものを関数 `Denny()` として定義。年齢別の l_x を返す
- データ (rlx) とモデル (mlx) の年齢ごとの差の二乗和を返す関数 `fDenny()` を定義
- `optim()` で `fDenny()` を最小化するパラメータ a, b, c を得る関数 `fitDenny()` を定義
- データをファイルから読み込み, 1 年分ずつ `fitDenny()` を実行してパラメータを行列に保存していく





Denny の年齢別死亡モデル

- fmsb パッケージに定義済み
- より単純な Denny のモデルを試してみる
 - ◆ `> library(fmsb)`
 - ◆ `> Denny`
 - ◆ `function (a, b, c, t)`
 - ◆ `{`
 - ◆ `lx <- ifelse(t < 105, 1/(1 + a * (t/(105 - t))^3 + b * sqrt(exp(t/(105 - t)) - 1) + c * (1 - exp(-2 * t))), 0)`
 - ◆ `lx <- ifelse(lx > 1, 1, lx)`
 - ◆ `return(lx * 1e+05)`
 - ◆ `}`
 - ◆ `<environment: namespace:fmsb>`

$$\ell(t) = \frac{1}{a \times \left(\frac{t}{105-t}\right)^3 + b \times \sqrt{\exp\left(\frac{t}{105-t}\right) - 1} + c \times (1 - \exp(-2t))}$$



Denny モデルの当てはめ

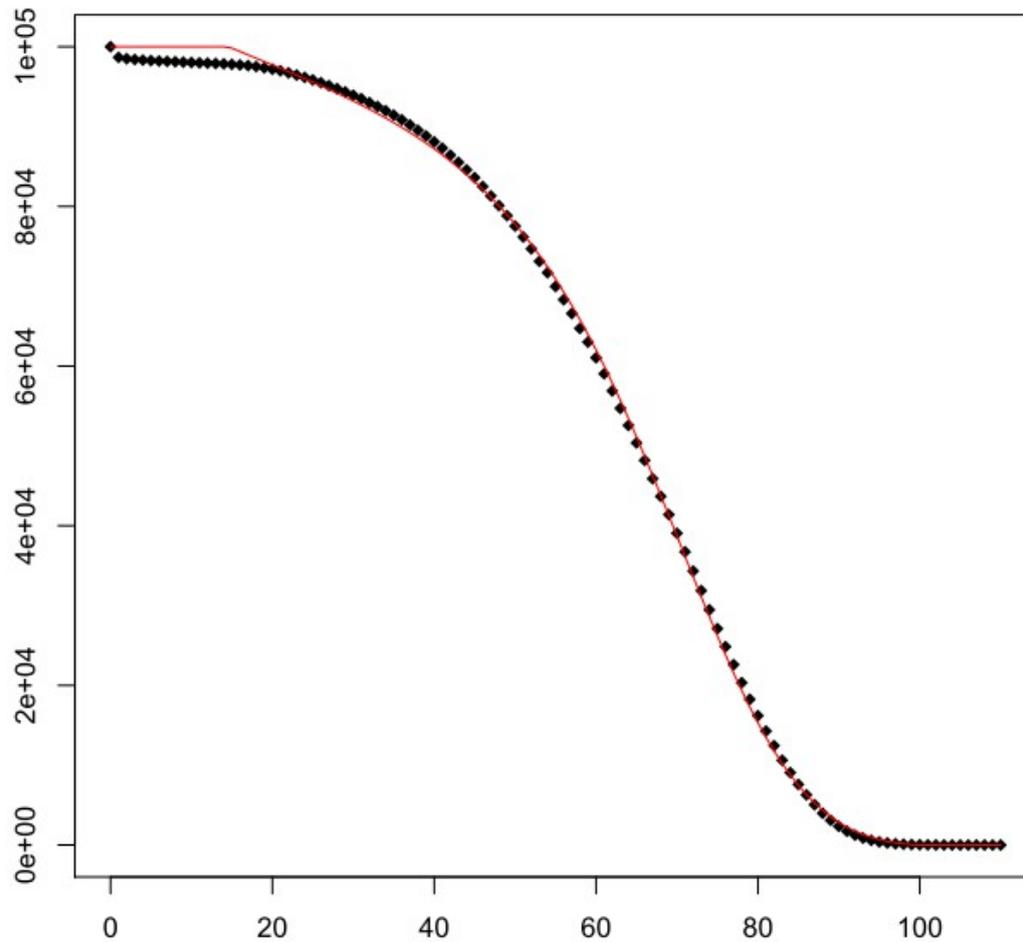
- `library(fmsb)`
- `ukm <- read.delim("./data/ukmdx.txt") # ウクライナ男性の dx の読込`
- `ryears <- 1961:2005`
- `fittedpar <- matrix(rep(0, 10*length(ryears)), length(ryears), 10)`
- `colnames(fittedpar) <- c("fa", "fb", "fc", "fRMSE", "fConv", "ma", "mb", "mc", "mRMSE", "mConv")`
- `fittedlxs <- matrix(rep(0, 111*length(ryears)), 111, length(ryears))`
- `colnames(fittedlxs) <- paste("Y", ryears, sep="")`
- `# calc for females`
- `for (yy in ryears) {`
- `lx <- dxtolx(subset(ukf, Year==yy)$dx)`
- `params <- rep(0,5)`
- `params <- fitDenny(, lx, Method="BFGS", mode=2)`
- `fittedpar[yy-1960, 1:5] <- params`
- `fittedlx <- Denny(params[1], params[2], params[3], 0:(length(lx)-1))`
- `fittedlxs[, paste("Y", yy, sep="")] <- fittedlx`
- `}`
- `write.table(fittedlxs, "./ukflx-model-denny.txt", quote=FALSE, sep="\t", row.names=FALSE) # ファイル保存`
- `# 女性も同様に推定`



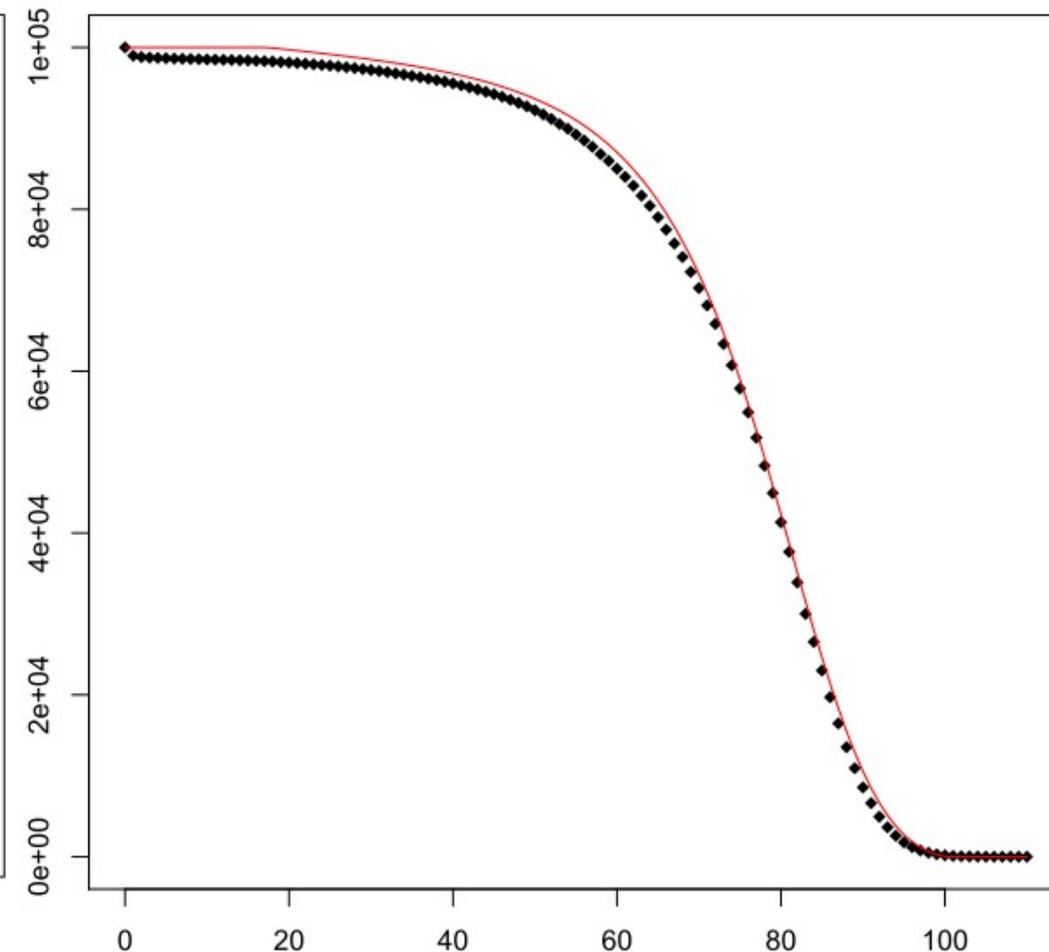
Denny のモデルを Ix に当てはめる

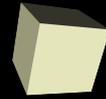
- 1961-2005 の間, 男女別々に→だいたい良い近似
- パラメータは forecast() で使うためにファイルに保存

Ix in Ukraine females in 2001



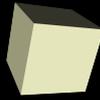
Ix in Ukraine females in 2001





時系列解析により係数を求め、信頼区間付きで予測

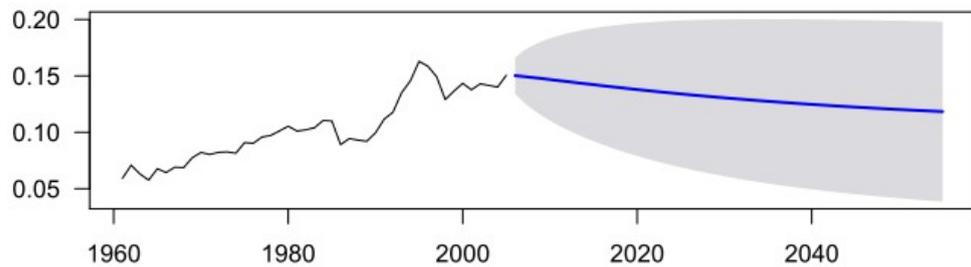
- `forecast` パッケージを用い、Arfima モデルを当てはめ、将来のパラメータをテキスト保存
- どのモデルを使うか、当てはめ方はどうすべきかは、まだ検討が不十分
 - `dennyparams <- read.delim("./ukraine-params-denny.txt")`
 - `library(forecast)`
 - `# females parameter a`
 - `ffa <- ts(dennyparams$fa, freq=1, start=1961)`
 - `fitffa <- arfima(ffa)`
 - `print(projffa <- forecast(fitffa, level=c(0.95), h=50))`
 - `# females parameter b`
 - `ffb <- ts(dennyparams$fb, freq=1, start=1961)`
 - `fitffb <- arfima(ffb)`
 - `print(projffb <- forecast(fitffb, level=c(0.95), h=50))`
 - `# females parameter c`
 - `ffc <- ts(dennyparams$fc, freq=1, start=1961)`
 - `fitffc <- arfima(ffc)`
 - `print(projffc <- forecast(fitffc, level=c(0.95), h=50))`
 - `write.table(data.frame(YEAR=2006:2055,PFA=projffa$mean, PFB=projffb$mean, PFC=projffc$mean),"projected-denny-females.txt",sep="\t",row.names=FALSE)`
 - `# 男性も同様に実行。`



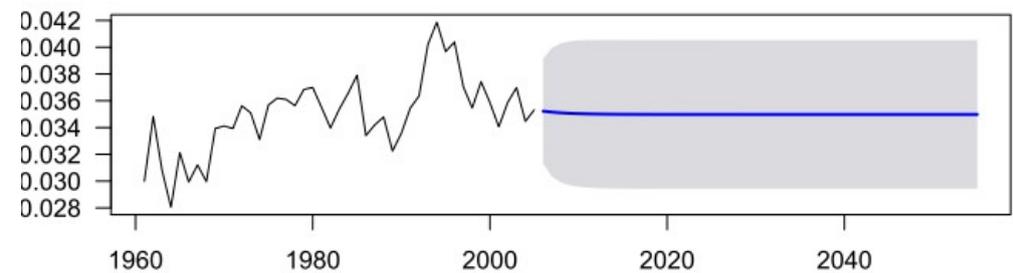
時系列解析により係数を求め、信頼区間付きで予測

- Denny モデルをウクライナの死亡データに当てはめて得られた係数 a , b , c について男女別に Arfima モデルを当てはめ、予測した結果 (左が男性, 右が女性)

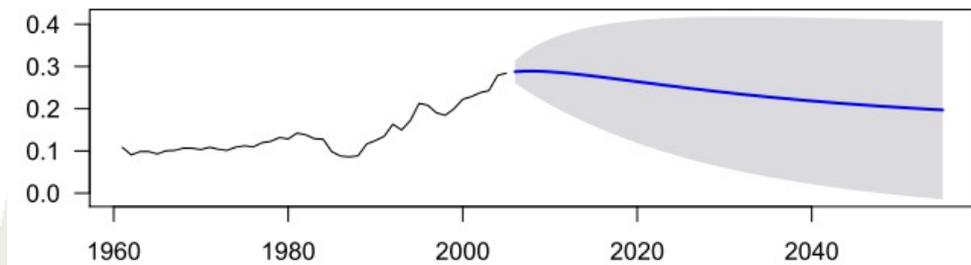
Forecasting parameter a of Denny's model fitted for Ukraine males



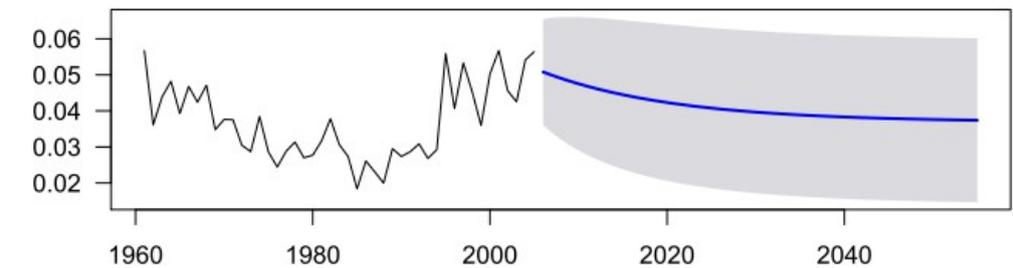
Forecasting parameter a of Denny's model fitted for Ukraine females



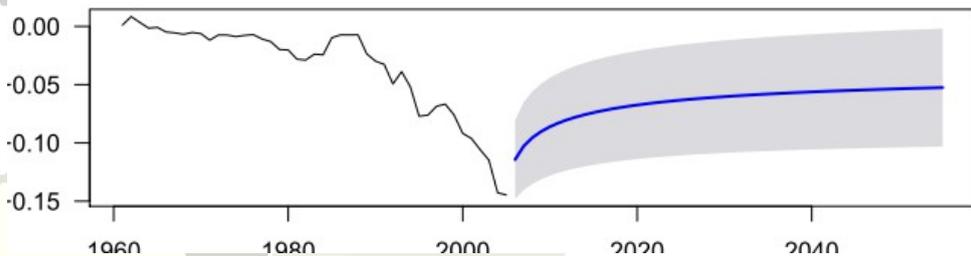
Forecasting parameter b of Denny's model fitted for Ukraine males



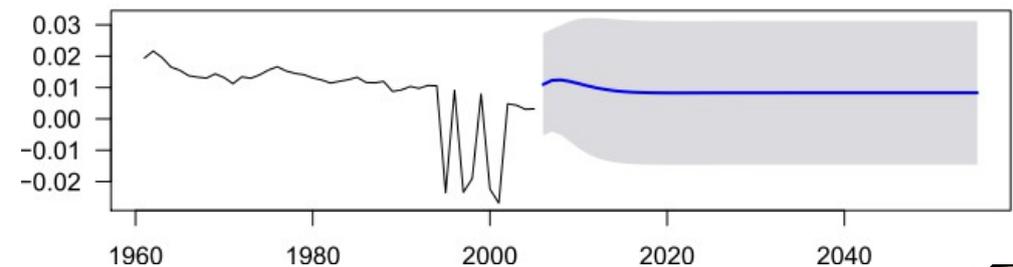
Forecasting parameter b of Denny's model fitted for Ukraine females



Forecasting parameter c of Denny's model fitted for Ukraine males



Forecasting parameter c of Denny's model fitted for Ukraine females



女性の年齢別出生率の単純なモデル化

- Hadwiger のモデル(下式)
- `>library(fmsb)`
- `>Hadwiger`
- `function (a = 3.4, b = 2.5, c = 22.2)`
- `{`
- `ages <- 15:54`
- `rval <- a * b/c * (c/ages)^1.5 * exp(-b^2 * (c/ages + ages/c - 2))`
- `names(rval) <- ages`
- `return(rval)`
- `}`
- `<environment: namespace:fmsb>`

$$\text{ASFR}(x) = (ab/c) \times (c/x)^{3/2} \times \exp(-b^2 \times (c/x + x/c - 2))$$



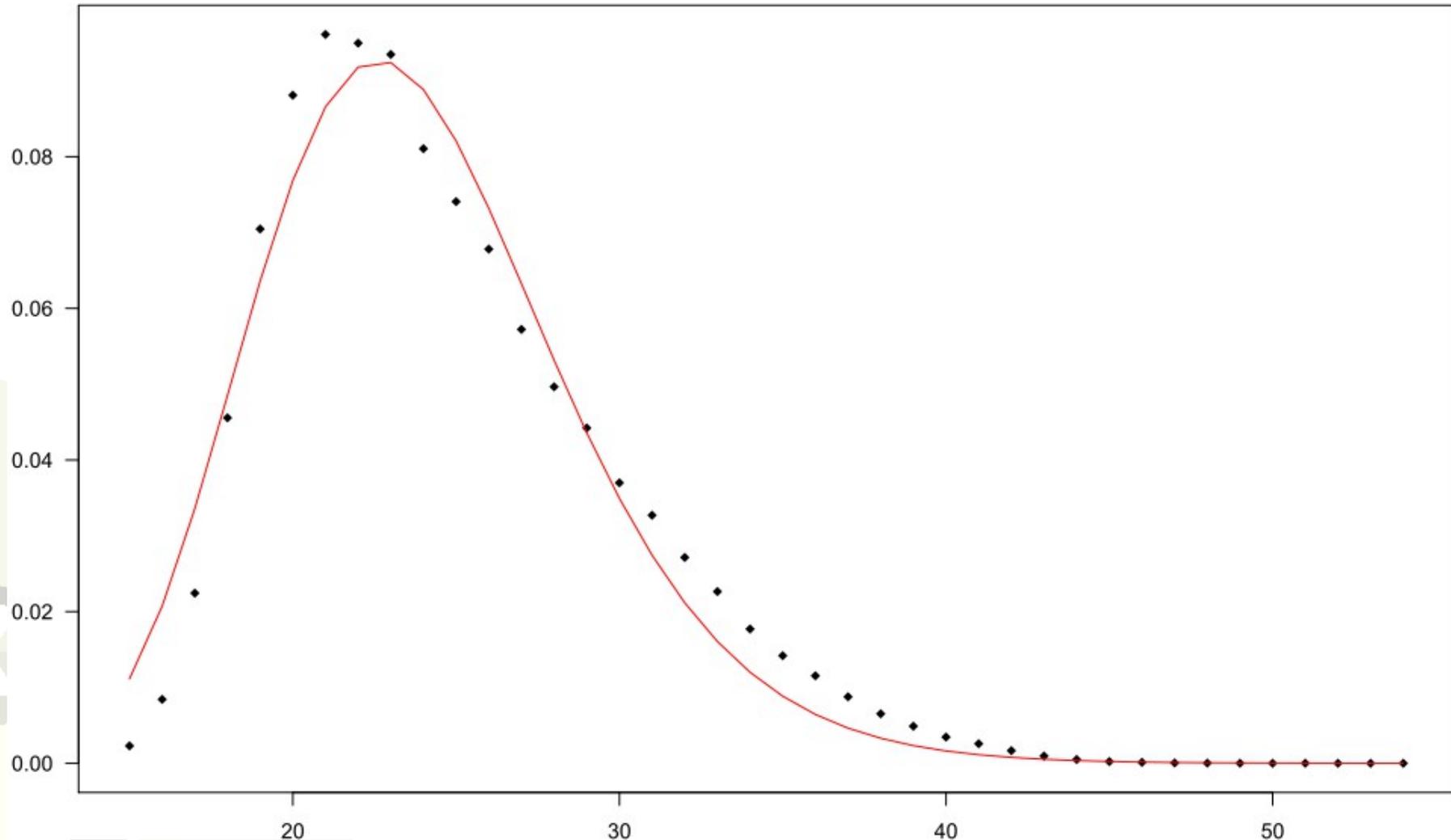
Hadwiger モデルの当てはめ

- `library(fmsb)`
- `asfr <- read.delim("./ASFR.txt")`
- `ryears <- 1961:2005`
- `ages <- 15:54`
- `fittedpar <- matrix(rep(0, 5*length(ryears)), length(ryears), 5)`
- `colnames(fittedpar) <- c("a", "b", "c", "RMSE", "Conv")`
- `fittedasfr <- matrix(rep(0, length(ages)*length(ryears)), length(ages), length(ryears))`
- `colnames(fittedasfr) <- paste("Y", ryears, sep="")`
- `# calc for females`
- `for (yy in ryears) {`
- `pasfr <- subset(asfr, ((Year==yy)&(Age %in% ages)))$ASFR`
- `params <- c(3.4, 2.5, 22.2, 0, 0)`
- `params <- fitHad(params[1:3], pasfr, Method="Nelder-Mead")`
- `fittedpar[yy-1960,] <- params`
- `fittedpasfr <- Hadwiger(params[1], params[2], params[3])`
- `fittedasfr[, paste("Y", yy, sep="")] <- fittedpasfr`
- `}`
- `write.table(fittedasfr, "./ukfittedasfr-hadwiger.txt", quote=FALSE, sep="\t", row.names=FALSE)`
- `write.table(fittedpar, "./ukasfr-params-hadwiger.txt", quote=FALSE, sep="\t", row.names=FALSE)`

女性の年齢別出生率の単純なモデル化

- 1961-2005 年の女性の年齢別出生率に当てはめた結果
- どの年もまあ当てはまる

ASFR in Ukraine females in 2001



Hadwiger モデルパラメータからの将来予測

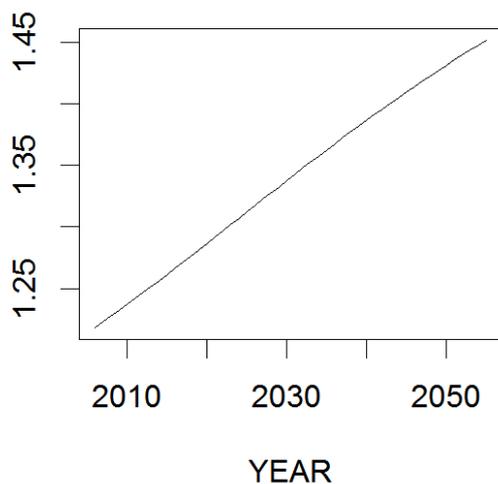
```
■ params <- read.delim("./Ukraine-fertility/ukasfr-params-hadwiger.txt")
■ library(forecast) # Rob Hyndman が開発した予測用パッケージ
■ library(fmsb)
■ # parameter a
■ ffa <- ts(params$a, freq=1, start=1961)
■ fitffa <- arfima(ffa)
■ print(projffa <- forecast(fitffa, level=c(0.95), h=50))
■ # parameter b
■ ffb <- ts(params$b, freq=1, start=1961)
■ fitffb <- arfima(ffb)
■ print(projffb <- forecast(fitffb, level=c(0.95), h=50))
■ # parameter c
■ ffc <- ts(params$c, freq=1, start=1961)
■ fitffc <- arfima(ffc)
■ print(projffc <- forecast(fitffc, level=c(0.95), h=50))
■ proja <- as.vector(projffa$mean)
■ projb <- as.vector(projffb$mean)
■ projc <- as.vector(projffc$mean)
■ TFRs <- rep(0, length(proja))
■ for (i in 1:length(proja)) {
■   TFRs[i] <- sum(Hadwiger(proja[i], projb[i], projc[i]))
■ }
■ write.table(data.frame(YEAR=2006:2055, TFRs, proja, projb, projc),"projected-
hadwiger-asfrs.txt",sep="\t",row.names=FALSE)
```



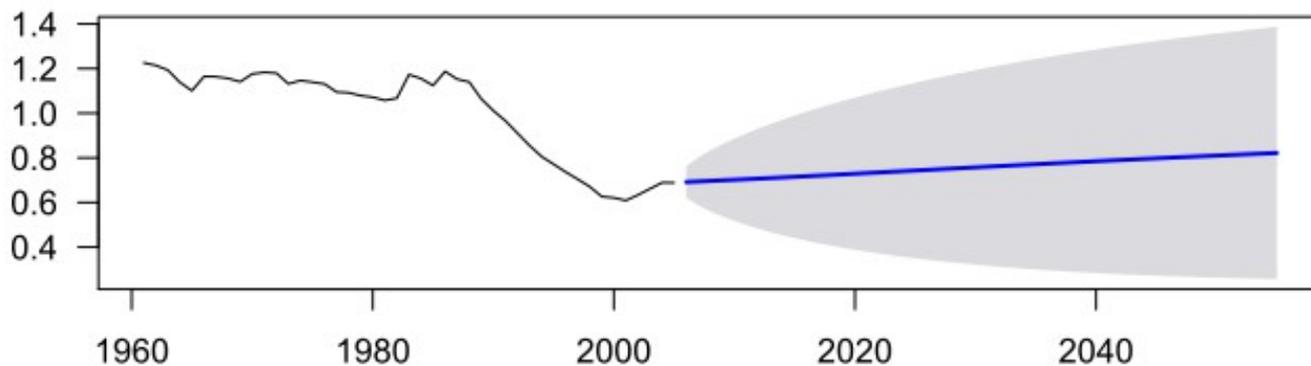
Hadwiger モデルパラメータの時系列解析将来予測

- Arfima モデル (要検討)
- forecast パッケージを使った当てはめ結果が右の図
- c がほぼ初婚または初産年齢に対応している

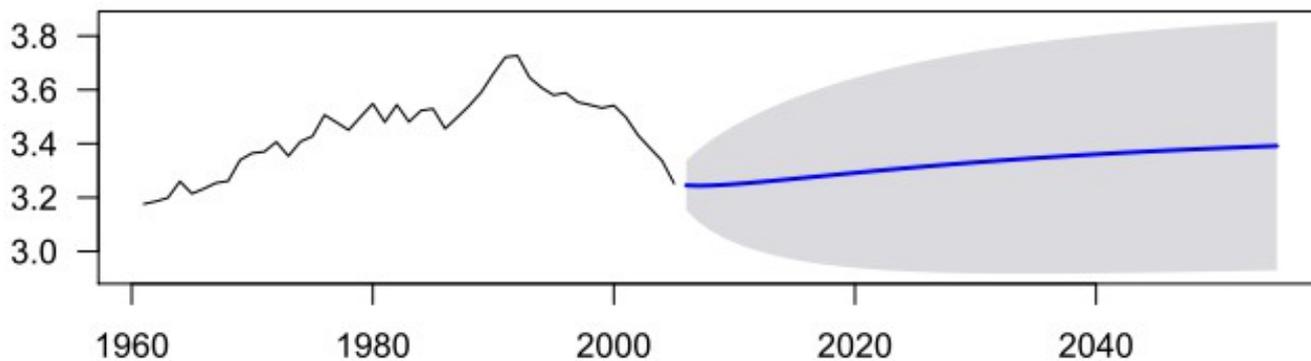
TFR予測



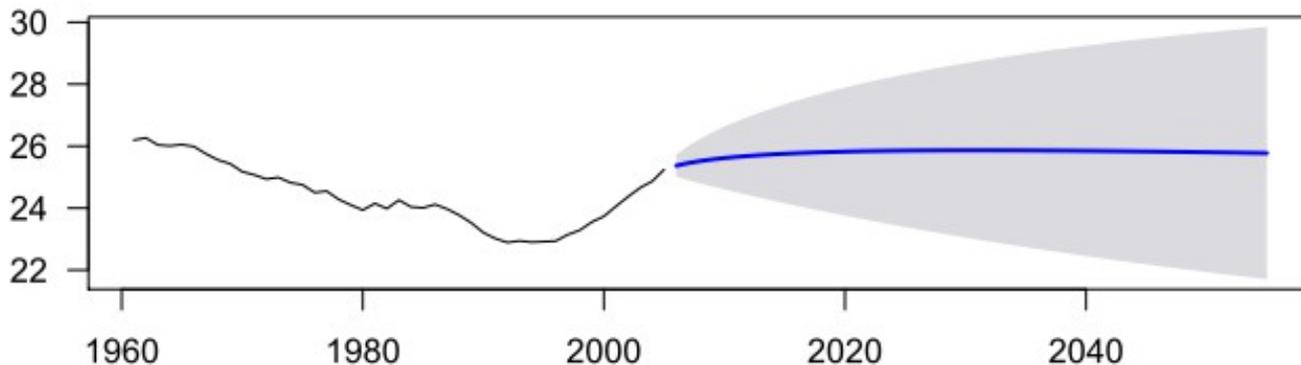
Forecasting parameter a of Hadwiger's model fitted for Ukraine females



Forecasting parameter b of Hadwiger's model fitted for Ukraine females



Forecasting parameter c of Hadwiger's model fitted for Ukraine females





保存しておいたパラメータをまとめて人口予測

- # 予測された人口は人口ピラミッドとして animation パッケージを使って動画にして表示
- `pop <- read.delim("./ukrainepop2001.txt")`
- `dfp <- read.delim("./projected-denny-females.txt")`
- `dmp <- read.delim("./projected-denny-males.txt")`
- `hp <- read.delim("./projected-hadwiger-asfrs.txt")`
- `library(pyramid); library(fmsb);`
- # Setting
- `wx <- 101; mp <- fp <- matrix(rep(0, wx*50), wx, 50) # For 50 years`
- `mp[,1] <- pop$Males; fp[,1] <- pop$Females`
- # Start simulation calculation
- `for (j in 2:50) {`
- `asfr <- c(rep(0, 14), Hadwiger(hp$proja[j-1], hp$projb[j-1], hp$projc[j-1]), rep(0, wx-54))`
- `mmx <- qxtomx(lxtoqx(Denny(dmp$PMA[j-1], dmp$PMB[j-1], dmp$PMC[j-1], 1:(wx-1))))`
- `fmx <- qxtomx(lxtoqx(Denny(dmp$PMA[j-1], dmp$PMB[j-1], dmp$PMC[j-1], 1:(wx-1))))`
- `baby <- sum(fp[, j-1]*asfr)`
- `babym <- as.integer(baby*1.06/2.06+0.5); mp[1, j] <- babym`
- `babyf <- as.integer(baby*1/2.06+0.5); fp[1, j] <- babyf`
- `mp[2:wx, j] <- as.integer(mp[1:(wx-1), j-1]*(1-mmx))`
- `fp[2:wx, j] <- as.integer(fp[1:(wx-1), j-1]*(1-fmx))`
- `}`
- # これは期待値だけなので、今後、95% 信頼区間の下限と上限の組合せを試す予定



保存しておいたパラメータをまとめて人口予測

- # Draw graphs
- library(animation) # animation パッケージを使う
- ani.options(interval = 0.5, outdir=getwd())
- saveVideo({ par(cex=0.8, las=1)
- pyramid(data.frame(M=pop\$Males/1000,
F=pop\$Females/1000, A=pop\$Age), Laxis=0:5*100,
AxisFM="d", Cstep=10)
- for (i in c(2:50)) { pyramid(data.frame(M=mp[, i]/1000,
F=fp[, i]/1000, A=pop\$Age),
■ Llab="Males\n(x1000)", Rlab="Females\n(x1000)",
Clab="",
■ Laxis=0:5*100, AxisFM="d", Cstep=10,
main=sprintf("Year %d projected", i+2005))
- ani.pause()
- }}, video.name="changing-pyramid2.mp4", other.opts="-b
300k")
- # mp4 で保存。RStudio から実行すると、ファイル生成後、
自動的に再生される

シミュレーション結果(これはアニメ GIF)

