Homework 03

Ihsan Kahveci

2022-04-25

Contents

1 Questions

1.1 *Q1*

Table 1: Peru female age-specific mortality rates, 1950-2020

age	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020
0	0.167	0.155	0.136	0.123	0.103	0.095	0.078	0.063	0.048	0.034	0.023	0.016	0.013	0.012	0.010
1	0.035	0.030	0.024	0.020	0.015	0.012	0.010	0.007	0.005	0.003	0.002	0.001	0.001	0.001	0.001
5	0.006	0.005	0.004	0.003	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.000
10	0.003	0.003	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000
15	0.004	0.004	0.003	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.000	0.000	0.000
20	0.006	0.006	0.005	0.004	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
25	0.007	0.006	0.005	0.005	0.004	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001
30	0.007	0.006	0.006	0.005	0.004	0.003	0.003	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.001
35	0.008	0.007	0.006	0.005	0.004	0.004	0.003	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.001
40	0.008	0.008	0.007	0.006	0.005	0.004	0.004	0.004	0.003	0.003	0.003	0.002	0.002	0.002	0.002
45	0.009	0.009	0.008	0.007	0.006	0.005	0.005	0.005	0.004	0.004	0.004	0.003	0.003	0.003	0.002
50	0.013	0.012	0.011	0.010	0.008	0.007	0.007	0.007	0.006	0.005	0.005	0.004	0.004	0.004	0.003
55	0.018	0.016	0.015	0.014	0.012	0.011	0.010	0.009	0.008	0.008	0.007	0.006	0.006	0.005	0.005
60	0.029	0.027	0.025	0.023	0.020	0.017	0.017	0.015	0.013	0.012	0.011	0.010	0.009	0.008	0.008
65	0.046	0.043	0.040	0.037	0.032	0.029	0.028	0.024	0.020	0.018	0.017	0.015	0.014	0.013	0.012
70	0.077	0.073	0.068	0.064	0.057	0.052	0.047	0.040	0.035	0.031	0.029	0.027	0.024	0.022	0.020
75	0.127	0.121	0.113	0.108	0.098	0.092	0.078	0.066	0.061	0.056	0.051	0.046	0.041	0.036	0.033
80	0.199	0.192	0.183	0.176	0.163	0.155	0.137	0.118	0.110	0.100	0.091	0.082	0.072	0.062	0.058
85	0.283	0.275	0.265	0.257	0.241	0.234	0.217	0.194	0.189	0.174	0.158	0.142	0.125	0.108	0.102
90	0.375	0.367	0.357	0.348	0.330	0.324	0.312	0.295	0.286	0.266	0.245	0.222	0.198	0.172	0.164
95	0.469	0.461	0.450	0.439	0.420	0.415	0.399	0.391	0.388	0.365	0.341	0.314	0.284	0.251	0.243
100	0.500	0.567	0.558	0.559	0.563	0.559	0.535	0.537	0.536	0.510	0.479	0.443	0.407	0.367	0.363

1.1.1 *Q1.a*

The *Lee-Carter* model is defined as:

$$log(m_{x,t}) = a_x + k_t b_x + \epsilon_{x,t}$$

One way to fit this model is the apply a set of constraints to b_x and k_t. Under the constraints $\sum b_x = 1$ and $\sum k_t = 0$, we get:

$$\hat{a}_x = \frac{1}{T}\sum_{t=1}^{T} log(m_{x,t})\hat{k}_t = \sum_{x=0}^{A-1}\left[log(m_{x,t}) - a_x\right]$$

With these constraints, we can fit a least squares regression on $\hat{b}_x$ with an intercept of 0. Below are the estimated parameters and mortality rates

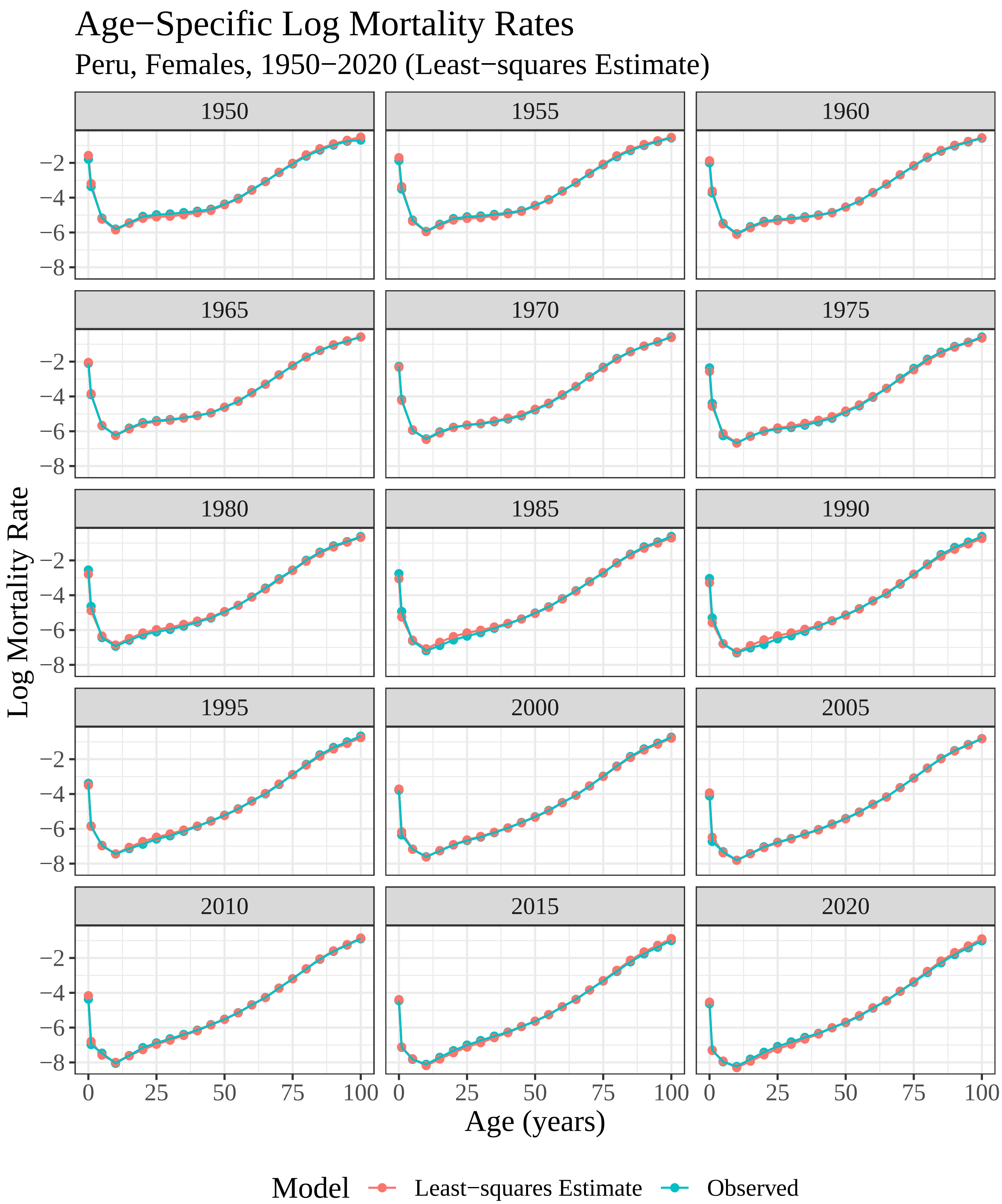
Note: This model was fit using the MortCast package.

Table 2: Lee-Carter model parameter estimates (least squares method)

Age	$\hat{a}_x$	$\hat{b}_x$
0	-3.031	0.079
1	-5.217	0.110
5	-6.557	0.072
10	-7.057	0.066
15	-6.682	0.065
20	-6.356	0.064
25	-6.147	0.057
30	-5.998	0.051
35	-5.808	0.045
40	-5.605	0.040
45	-5.360	0.034
50	-5.039	0.034
55	-4.680	0.033
60	-4.207	0.034
65	-3.753	0.037
70	-3.212	0.037
75	-2.681	0.036
80	-2.144	0.033
85	-1.671	0.026
90	-1.291	0.021
95	-1.000	0.016
100	-0.706	0.010

Table 3: Lee-Carter model parameter estimates (least squares method)

Year	$\hat{k}_t$
1950	18.369
1955	16.806
1960	14.521
1965	12.445
1970	8.938
1975	5.916
1980	2.995
1985	-0.301
1990	-3.255
1995	-5.839
2000	-8.640
2005	-11.450
2010	-14.317
2015	-17.167
2020	-19.022



Age−Specific Log Mortality Rates
Peru, Females, 1950−2020 (Least−squares Estimate)
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
Log Mortality Rate
−2
−4
−6
−8
0
25
50
75
100
Age (years)
Model
Least−squares Estimate
Observed

1.1.2 *Q1.b*

A singular value decomposition can also be used to implement the *Lee-Carter* model, where the matrix $C = \left[log(m_{x,t}) - a_x\right]$ is used to get b_x and k_t with the decomposition:

$$\text{SVD}(C) = U\Lambda V^{\top}$$

Where:

$$b_x = U_{x,1} \div \sum_x U_{x,1} k_t = \left[V^{\top}\right]_{1,t} \times \sum U_{x,1} \times \Lambda_1$$

are the normalized estimates. The estimated parameters and fitted values are below.

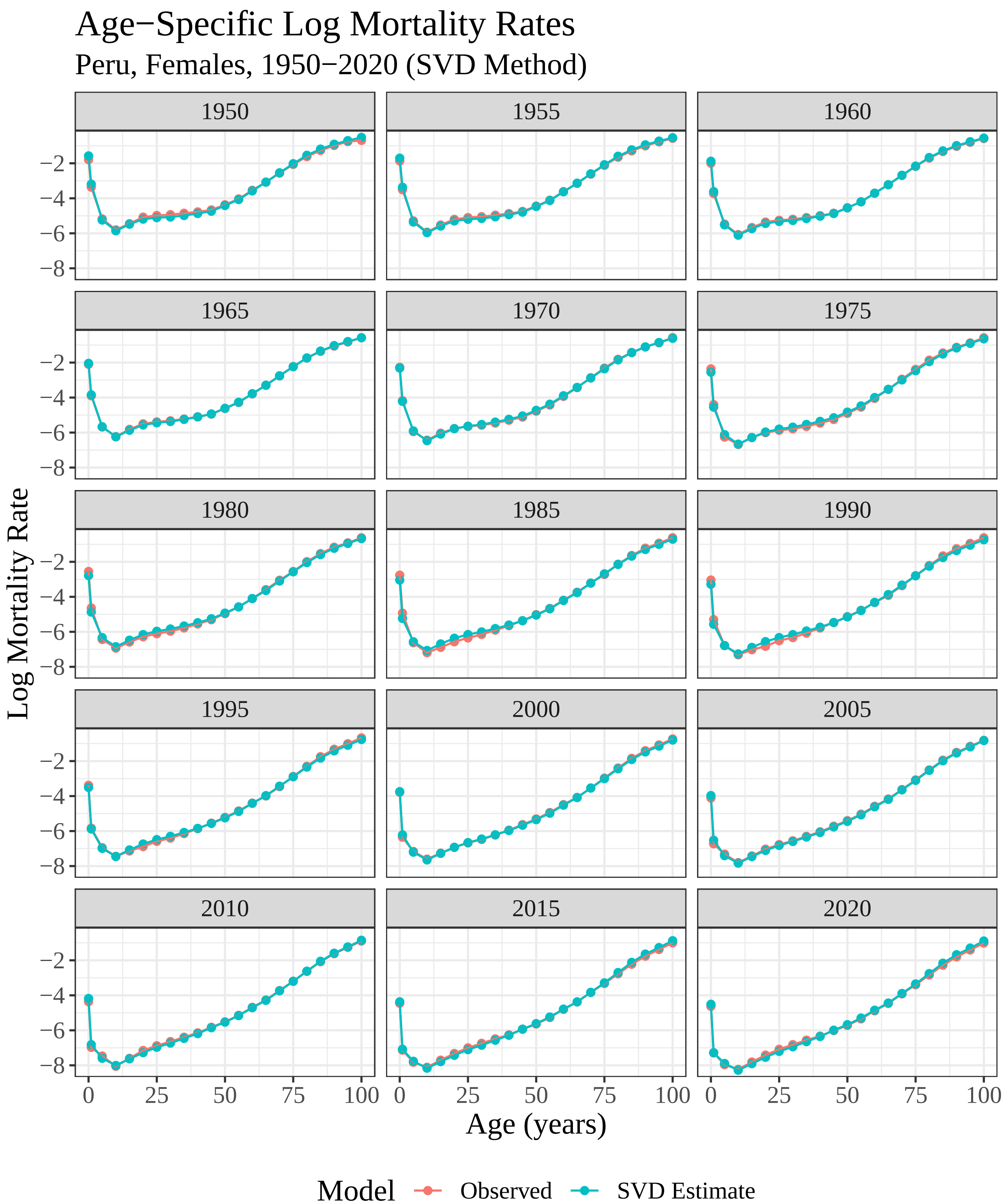
Table 4: Lee-Carter model parameter estimates (SVD method)

Age	b_x
0	0.079
1	0.110
5	0.072
10	0.066
15	0.066
20	0.064
25	0.057
30	0.051
35	0.045
40	0.040
45	0.034
50	0.034
55	0.033
60	0.034
65	0.037
70	0.037
75	0.036
80	0.033
85	0.026
90	0.021
95	0.016
100	0.010

Table 5: Lee-Carter model parameter estimates (SVD method)

Year	k_t
1950	18.342
1955	16.739
1960	14.484
1965	12.431
1970	9.056
1975	6.071
1980	3.136
1985	-0.151

Year	k_t
1990	-3.186
1995	-6.015
2000	-9.037
2005	-11.841
2010	-14.490
2015	-16.879
2020	-18.658



Comparing the estimated M_x from the normalized SVD to our observed data, we see the estimates also track the observed data well. Note if we were to use the unnormalized SVD, the estimated M_x would remain mostly constant over time, matching the middle years the closest (since those years are closest to the mean over time).

1.1.3 *Q1.c*

Table 6: Lee-Carter model comparison

Age	LS $\hat{b}_x$	SVD $\hat{b}_x$	diff
0	0.07923	0.07933	0.00009
1	0.11030	0.11046	0.00016
5	0.07178	0.07172	0.00006
10	0.06556	0.06556	0.00001
15	0.06549	0.06551	0.00002
20	0.06354	0.06354	0.00000
25	0.05675	0.05676	0.00001
30	0.05052	0.05053	0.00001
35	0.04497	0.04497	0.00000
40	0.04020	0.04018	0.00002
45	0.03397	0.03395	0.00003
50	0.03421	0.03419	0.00003
55	0.03318	0.03316	0.00002
60	0.03444	0.03443	0.00002
65	0.03679	0.03680	0.00000
70	0.03658	0.03659	0.00001
75	0.03590	0.03589	0.00001
80	0.03286	0.03285	0.00001
85	0.02643	0.02640	0.00003
90	0.02085	0.02082	0.00004
95	0.01618	0.01614	0.00004
100	0.01025	0.01023	0.00003

Table 7: Lee-Carter model comparison

Year	LS $\hat{k}_t$	SVD $\hat{k}_t$	diff
1950	18.369	18.342	0.027
1955	16.806	16.739	0.067
1960	14.521	14.484	0.037
1965	12.445	12.431	0.014
1970	8.938	9.056	0.118
1975	5.916	6.071	0.154
1980	2.995	3.136	0.141
1985	-0.301	-0.151	0.150
1990	-3.255	-3.186	0.069
1995	-5.839	-6.015	0.176
2000	-8.640	-9.037	0.397
2005	-11.450	-11.841	0.391
2010	-14.317	-14.490	0.172
2015	-17.167	-16.879	0.288
2020	-19.022	-18.658	0.363

Table 8: $RMSE$ of estimated $log(M_x)$ against observed $log(M_x)$

Model	RMSE
Least-squares Estimate	4.7202
SVD Estimate	4.7203

Looking at the previous two plots, and comparing the root mean-squared-error of the log-transformed fitted M_x against the observed M_x, we see , that both methods performs almost the same.

1.1.4 *Q1.d*

The *Lee-Carter* method can also be used to obtain a probabilistic forecast of mortality. Here we use the *demography R* package to forecast mortality for females in Peru in the period 2020-2025. Below are the predicted mortality index k_t and mortality rate for females in the 75-80 age group. The confidence level is set to 95%.

Table 9: Forecast k_t

Est. k_t	95% Low	95% High
-2.4	-3.564	-1.236

Table 10: Forecast mortality for age group 75-80

Est. M_x	95% Low	95% High
0.03	0.029	0.032

1.2 *Q2*

1.2.1 *Q2.a*

We can generally define the posterior distribution as the product of a likelihood and prior function:

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$
$$P(\theta \mid X = x) \propto P(X = x \mid \theta) \times P(\theta)$$

where, given the parameters of the problem, the likelihood takes the form of a binomial distribution, and the prior probability follows a uniform distribution (or, a beta distribution with $\alpha = \beta = 1$):

$$X \mid \theta \sim \text{Binomial}(X, \theta)$$
$$\theta \sim \text{Beta}(1, 1)$$

Binomial likelihood function is defined as k successes in n trials, each with a probability p of occurring:

$$\text{Likelihood: } P(X = k \mid \theta; n) = \binom{n}{k}\theta^k(1-\theta)^{n-k}$$

Prior, θ, is a beta distribution, it acts as a conjugate prior for the binomial likelihood function. Hence the posterior distribution is a beta distribution in the form:

$$\text{Posterior: } P(\theta \mid X = k; \alpha, \beta) = \text{Beta}(\alpha + k, n - k + \beta)$$

From the problem, we know $k = 43$ divorces occurred over the period 2005-2015 from a sample of $n = 112$ married people in 2005. This leads us to our final analytic posterior distribution:

$$\begin{aligned} P(\theta \mid X = k; \alpha, \beta, n) &= \text{Beta}(\alpha + k, n - k + \beta) \\ P(\theta \mid X = 43; 1, 1, 112) &= \text{Beta}(1 + 43, 112 - 43 + 1) \\ &= \text{Beta}(44, 70) \end{aligned}$$

1.2.2 *Q2.b*

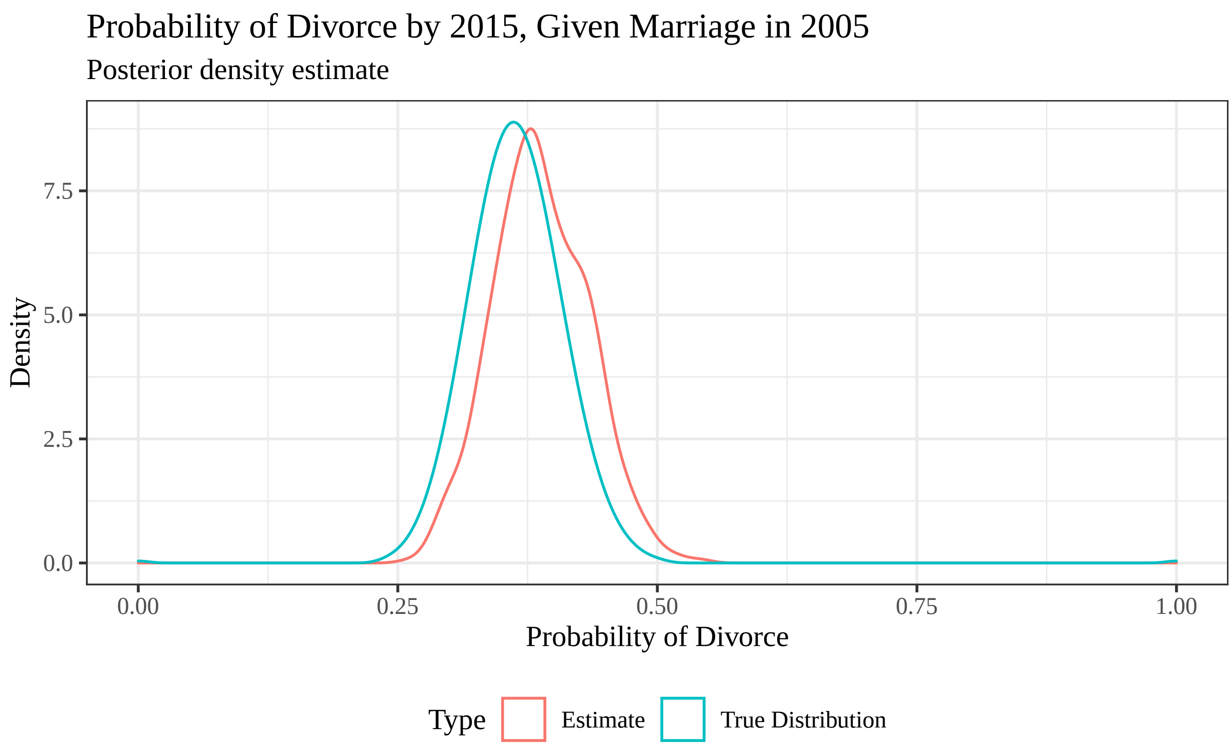
From our analytic posterior distribution of θ, we simulate a new sample of $1,000$ people. From this sample, we can get a 95% Bayesian confidence interval for θ by pulling the 2.5^{th} and 97.5^{th} percentile of the data.

Table 11: Posterior distribution summary

Mean	Median	95% Low	95% High
0.388	0.384	0.3	0.48

1.2.3 *Q2.c*

With the new sample from our posterior, we can plot a nonparametric density estimate for our posterior. The true distribution is also included for reference. Both density curves use a Gaussian kernel.


Probability of Divorce by 2015, Given Marriage in 2005
Posterior density estimate
7.5
5.0
2.5
0.0
Density
0.00
0.25
0.50
0.75
1.00
Probability of Divorce
Type
Estimate
True Distribution

1.3 *Q3*

In this scenario, we take both our likelihood and prior distributions to be normal:

$$X \mid \theta \propto N(\theta, \sigma^2)$$
$$\theta \propto N(\mu, \tau^2)$$

where we are given $\mu = 10$, $\tau = 3$, $\sigma = 4$, and $x = 9.46$, our observed mean.

By Bayes theorem:

$$\Pr(\mu \,|\, \mathbf{x}, \sigma^2) \propto \Pr(\mathbf{x} \,|\, \mu, \sigma^2) \Pr(\mu)$$

$$\mathrm{N}(\mu_1, \tau_1^2) = \mathrm{N}(\mu, \sigma^2)\mathrm{N}(\mu_0, \tau_0^2)$$

where the posterior mean:

$$\mu_1 = \frac{\frac{\mu_0}{\tau_0^2} + \frac{n\bar{x}}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$$

and the posterior variance:

$$\tau_1^2 = (\frac{1}{\tau_0^2} + \frac{n}{\sigma^2})^{-1}$$

We can draw $1,000$ new values from this posterior distribution to get a new sample. From this sample, we can get a 95% Bayesian confidence interval for θ by pulling the 2.5^{th} and 97.5^{th} percentile of the data.

Table 12: Posterior distribution summary

Mean	Median	95% Low	95% High
9.445	9.472	7.914	10.826

2 Appendix

```r
# Prep work ---------------------------------------------------------------

# Load libraries
library(tidyverse)
library(MortalityLaws)
library(demogR)
library(MortCast)
library(demography)

# Helper functions
rmse <- function(x, y) sqrt(mean((x - y)^2))
mx_to_qx <- function(mx, n) 1 - exp(-1 * n * mx)
qx_to_mx <- function(qx, n) -1 * log(1 - qx) / n

# plotting posterior density estimates
post_draws_density <- function(data, ...) {
  ggplot(data, aes(...)) +
    geom_density() +
    theme_bw() +
    theme(
      text = element_text(family = "serif"),
      legend.position = "bottom"
    ) +
    labs(
      subtitle = paste0("Posterior density estimate"),
      y = "Density",
      color = "Type"
    )

}

# plotting log mortality models
plot_log_mort <- function(data, ...) {
  ggplot(data, aes(...)) +
    geom_point() +
    geom_line() +
    theme_bw() +
    theme(
      text = element_text(family = "serif", size = 15),
      legend.position = "bottom"
    ) +
    labs(
      title = "Age-Specific Log Mortality Rates",
      subtitle = "Peru, Females, 2020-2025",
      x = "Age (years)",
      y = "Log Mortality Rate",
      color = "Model"
    )
}

# Control randomness
```

```r
set.seed(57)


# Load data
data(mxF, package = "wpp2019")
peru_mort <- mxF %>%
  filter(name == "Peru") %>%
  select(-country_code, -name) %>%
  pivot_longer(-age, names_to = "period", values_to = "Mx") %>%
  extract(period, "year", regex = "(^[0-9]{4})", convert = TRUE) %>%
  filter(year < 2025) %>%
  select(year, age, everything()) %>%
  arrange(year, age)


# Question 1 -------------------------------------------------------------

peru_mort_wide <- peru_mort %>% pivot_wider(names_from = year, values_from = Mx)
peru_mort_mat <- peru_mort_wide %>% column_to_rownames("age") %>% as.matrix()

# Question 1a ------------------------------------------------------------

knitr::kable(
  peru_mort_wide,
  booktabs = TRUE,
  digits = 3,
  caption = "Peru female age-specific mortality rates, 1950-2020"
)

# Question 1a ------------------------------------------------------------

model_LC <- leecarter.estimate(peru_mort_mat)

LC_lsq_mx <- with(model_LC, apply(
  matrix(bx) %*% matrix(kt, nrow = 1), 2, function(x) ax + x
))
dimnames(LC_lsq_mx) <- dimnames(peru_mort_mat)

peru_mort_lsq <- LC_lsq_mx %>%
  as_tibble(rownames = "age") %>%
  pivot_longer(-age, names_to = "year", values_to = "log_Mx") %>%
  mutate(
    year = as.integer(year),
    age = as.integer(age),
    `Least-squares Estimate` = exp(log_Mx),
  ) %>%
  select(-log_Mx)

peru_LC_compare <- peru_mort %>%
  left_join(peru_mort_lsq, by = c("age", "year")) %>%
  rename(Observed = Mx)

model_LC_ls_x_tbl <- tibble(
```

```r
  age = names(model_LC$ax),
  ax = model_LC$ax,
  least_squares_bx = model_LC$bx
)

model_LC_ls_t_tbl <- tibble(
  year = names(model_LC$kt),
  least_squares_kt = model_LC$kt
)

knitr::kable(
  model_LC_ls_x_tbl,
  booktabs = TRUE,
  digits = 3,
  col.names = c("Age", "$\\hat{a}_x$", "$\\hat{b}_x$"),
  eval = FALSE,
  caption = "Lee-Carter model parameter estimates (least squares method)"
)

knitr::kable(
  model_LC_ls_t_tbl,
  booktabs = TRUE,
  digits = 3,
  col.names = c("Year", "$\\hat{k}_t$"),
  eval = FALSE,
  caption = "Lee-Carter model parameter estimates (least squares method)"
)

peru_LC_compare %>%
  pivot_longer(!c(year, age), names_to = "model", values_to = "Mx") %>%
  filter(model %in% c("Observed", "Least-squares Estimate")) %>%
  plot_log_mort(x = age, y = log(Mx), color = model) +
  facet_wrap(vars(year), ncol = 3) +
  labs(subtitle = "Peru, Females, 1950-2020 (Least-squares Estimate)")

# Question 1b -------------------------------------------------------------

ax_hat <- rowMeans(log(peru_mort_mat))
model_LC_svd <- svd(apply(log(peru_mort_mat), 2, function(x) x - ax_hat), 1, 1)

# Get normalized bx and kt
LC_svd_bx <- model_LC_svd$u / sum(model_LC_svd$u)
LC_svd_kt <- t(model_LC_svd$v) * sum(model_LC_svd$u) * model_LC_svd$d[1]

LC_svd_mx <- apply(LC_svd_bx %*% LC_svd_kt, 2, function(x) ax_hat + x)
dimnames(LC_svd_mx) <- dimnames(peru_mort_mat)

peru_mort_svd <- LC_svd_mx %>%
  as_tibble(rownames = "age") %>%
  pivot_longer(-age, names_to = "year", values_to = "log_Mx") %>%
  mutate(
    year = as.integer(year),
    age = as.integer(age),
```

```r
    `SVD Estimate` = exp(log_Mx),
  ) %>%
  select(-log_Mx)

peru_LC_compare <- peru_LC_compare %>%
  left_join(peru_mort_svd, by = c("age", "year"))

model_LC_svd_x_tbl <- tibble(
  age = names(model_LC$bx),
  svd_bx = as.vector(LC_svd_bx)
)

model_LC_svd_t_tbl <- tibble(
  year = names(model_LC$kt),
  svd_kt = as.vector(LC_svd_kt),
)

knitr::kable(
  model_LC_svd_x_tbl,
  booktabs = TRUE,
  digits = 3,
  col.names = c("Age", "$b_x$"),
  eval = FALSE,
  caption = "Lee-Carter model parameter estimates (SVD method)"
)

knitr::kable(
  model_LC_svd_t_tbl,
  booktabs = TRUE,
  digits = 3,
  col.names = c("Year", "$k_t$"),
  eval = FALSE,
  caption = "Lee-Carter model parameter estimates (SVD method)"
)

peru_LC_compare %>%
  pivot_longer(!c(year, age), names_to = "model", values_to = "Mx") %>%
  filter(model %in% c("Observed", "SVD Estimate")) %>%
  plot_log_mort(x = age, y = log(Mx), color = model) +
  facet_wrap(vars(year), ncol = 3) +
  labs(subtitle = "Peru, Females, 1950-2020 (SVD Method)")

# Question 1c ----------------------------------------------------------

LC_compare_bx_tbl <- model_LC_ls_x_tbl %>%
  left_join(model_LC_svd_x_tbl, by = "age") %>%
  select(-ax) %>%
  mutate(abs_diff = abs(least_squares_bx - svd_bx))

LC_compare_kt_tbl <- model_LC_ls_t_tbl %>%
  left_join(model_LC_svd_t_tbl, by = "year") %>%
  mutate(abs_diff = abs(least_squares_kt - svd_kt))
```

```r
peru_LC_compare_fit <- peru_LC_compare %>%
  group_by(year) %>%
  select(-age) %>%
  mutate_at(vars(-group_cols()), log) %>%
  summarise_all(~rmse(., Observed)) %>%
  select(-Observed) %>%
  ungroup() %>%
  select(-year) %>%
  summarize_all(mean) %>%
  pivot_longer(everything(), names_to = "Model", values_to = "RMSE") %>%
  arrange(RMSE)

knitr::kable(
  LC_compare_bx_tbl,
  booktabs = TRUE,
  digits = 5,
  col.names = c("Age", "LS $\\hat{b}_x$", "SVD $\\hat{b}_x$", "diff"),
  eval = FALSE,
  caption = "Lee-Carter model comparison"
)

knitr::kable(
  LC_compare_kt_tbl,
  booktabs = TRUE,
  digits = 3,
  col.names = c("Year", "LS $\\hat{k}_t$", "SVD $\\hat{k}_t$", "diff"),
  eval = FALSE,
  caption = "Lee-Carter model comparison"
)

knitr::kable(
peru_LC_compare_fit,
booktabs = TRUE,
digits = 4,
caption = "$RMSE$ of estimated $log(M_x)$ against observed $log(M_x)$"
)

# Question 1d -------------------------------------------------------------

data(popF, package = "wpp2019")

mex_pop <- popF %>%
  filter(name == "Peru") %>%
  select(-country_code, -name ) %>%
  extract(age, "age", convert = TRUE) %>%
  column_to_rownames("age") %>%
  as.matrix() %>%
  `*`(1000)

mex_mort_tbl <- mxF %>%
  filter(name == "Peru") %>%
  select(-country_code, -name) %>%
  pivot_longer(-age, names_to = "period", values_to = "Mx") %>%
```

```r
  extract(period, "year", regex = "(^[0-9]{4})", convert = TRUE) %>%
  filter(year < 2025)

# Collapse 0 and 1-4 age group Mx
mex_mort_tbl_u5 <- mex_mort_tbl %>%
  filter(age < 5) %>%
  mutate(
    n = age * 3 + 1,
    sx = 1 - mx_to_qx(Mx, n)
  ) %>%
  group_by(year) %>%
  summarise(sx = prod(sx)) %>%
  mutate(
    age = 0,
    Mx = qx_to_mx(1 - sx, 5)
  ) %>%
  select(age, year, Mx)

mex_mort <- mex_mort_tbl %>%
  filter(age >= 5) %>%
  bind_rows(mex_mort_tbl_u5) %>%
  pivot_wider(names_from = year, values_from = Mx) %>%
  arrange(age) %>%
  column_to_rownames("age") %>%
  as.matrix()

demog_data <- demogdata(
  data = mex_mort,
  pop = mex_pop,
  ages = seq(0, 100, 5),
  years = seq(1950, 2020, 5),
  type = "mortality",
  label = "Peru",
  name = "Female"
)

model_mex_LC <- demography::lca(demog_data)

forecast_mex <- forecast(model_mex_LC, h = 1, level = 95, se = "innovonly")

forecast_mex_kt <- as_tibble(forecast_mex[["kt.f"]])

forecast_mex_mx <- forecast_mex$rate %>%
  as_tibble() %>%
  mutate_all(as.vector) %>%
  mutate(age = rownames(forecast_mex$rate$Female)) %>%
  select(age, Mx = Female, lower_95 = lower, upper_95 = upper)

forecast_mex_mx_75 <- forecast_mex_mx %>% filter(age == 75)

knitr::kable(
  forecast_mex_kt,
  booktabs = TRUE,
```

```r
  digits = 3,
  eval = FALSE,
  col.names = c("Est. $k_t$", "95% Low", "95% High"),
  caption = "Forecast $k_t$"
)

knitr::kable(
  select(forecast_mex_mx_75, -age),
  booktabs = TRUE,
  digits = 3,
  eval = FALSE,
  col.names = c("Est. $M_x$", "95% Low", "95% High"),
  caption = "Forecast mortality for age group 75-80"
)

# Question 2 -------------------------------------------------------------

prior_a <- 1
prior_b <- 1

n_married <- 112
divorced_n_obs <- 43


# Question 2a ------------------------------------------------------------

divorced_posterior <- function(n) {
  rbeta(n, prior_a + divorced_n_obs, n_married - divorced_n_obs + prior_b)
}


# Question 2b ------------------------------------------------------------

divorced_n_sims <- 1000

divorced_posterior_draws <- divorced_posterior(divorced_n_sims)

divorced_post_tbl <- tibble(
  Mean = mean(divorced_posterior_draws),
  Median = quantile(divorced_posterior_draws, .5),
  `95% Low` = quantile(divorced_posterior_draws, .025),
  `95% High` = quantile(divorced_posterior_draws, .975)
)

knitr::kable(
  divorced_post_tbl,
  booktabs = TRUE,
  digits = 3,
  caption = "Posterior distribution summary"
)

# Question 2c ------------------------------------------------------------
```

```
divorced_true_post <- tibble(
  value = qbeta(seq(0, 1, .001), 44, 77),
  type = "True Distribution"
)

divorced_est_post <- tibble(
  value = divorced_posterior_draws,
  type = "Estimate"
)

divorced_post_compare_tbl <- dplyr::bind_rows(
  divorced_true_post, divorced_est_post
)

divorced_post_plot <-
  post_draws_density(divorced_post_compare_tbl, x = value, color = type) +
  labs(
    title = "Probability of Divorce by 2015, Given Marriage in 2005",
    x = "Probability of Divorce")

divorced_post_plot
# Question 3 --------------------------------------------------------------

study_obs <- c(
   2.1, 9.8, 13.9, 11.3, 8.9, 15.7, 16.4, 4.5, 8.9, 11.9, 12.5, 11.1, 11.6,
  14.5, 9.6,  7.4,  3.3, 9.1,  9.4,  6.0, 7.4, 8.5,  1.6, 11.4,  9.7)

true_sd <- 4
prior_mean <- 10
prior_sd <- 3

n = length(study_obs)
data_mean = mean(study_obs)
study_post_mean = ((prior_mean/prior_sd^2) + ((n * data_mean)/true_sd^2))/((1/prior_sd^2) + (n/true_sd^
study_post_variance = 1/((1/prior_sd^2) + (n/true_sd^2))
study_post_sd = sqrt(study_post_variance)


study_posterior <- function(n) rnorm(n, study_post_mean, study_post_sd)

study_n_sims <- 1000

study_posterior_draws <- study_posterior(study_n_sims)

study_post_tbl <- tibble(
  Mean = mean(study_posterior_draws),
  Median = quantile(study_posterior_draws, .5),
  `95% Low` = quantile(study_posterior_draws, .025),
  `95% High` = quantile(study_posterior_draws, .975)
)


knitr::kable(
```

```
  study_post_tbl,
  booktabs = TRUE,
  digits = 3,
  caption = "Posterior distribution summary"
)
```