ESTIMATION OF DISTRIBUTION FOR LIVING TIME FOR CENSORED DATA

Tserenbat.O, PhD[1]; Gereltuya.T, MD[2]; Narangoo.G, MD[1]
[1]National University of Mongolia, [2]Mongolian University of Science and Technology

Abstract

In this study, to construct a lifetime distribution function, we used censored data that included a countries population pyramid and a life table.

We use a third-order B-spline function to construct the logarithm for force of mortality of living time.
Evaluation of B-spline parameters estimated by maximum likelihood estimation tested with criteria of the modified chi-squared goodness of the fit statistic.

Based on demographic data of the America (2019), Japan (2019), Germany (2019), Russia (2014),and Mongolia(2019) we calculate the estimation and criteria using Matlab.

Introduction

In any nation, setting up the force of mortality and the living time distribution are crucially important in actuarial science, life insurance, health, and demographic surveillance system. Thus, estimation of the force of mortality has become the most important issues among the researcher.
Let's suppose that X is the lifetime random variable.
Then four functions define and explain the characteristics of the distribution of X. These are the probability density function $f(x)$, the cumulative distribution function of lifetime $F(x) = P(X < x)$, the survival function $s(x) = 1 - P(X < x)$ and the force of mortality $\lambda(x)$. If one of the above four functions is known, the other three can be determined.
O.Tserenbat, T.Gereltuya and D.Nyamsuren constructed the approximate function living time distribution and the force of mortality.
They estimate the logarithm of the force of mortality in the form of the third order B-spline.
Its parameters have been estimated by the maximum likelihood estimation.
They examine the fit of the obtained distribution with the empirical distribution by using sequential procedure for the modified χ^2 goodness of fit statistics.

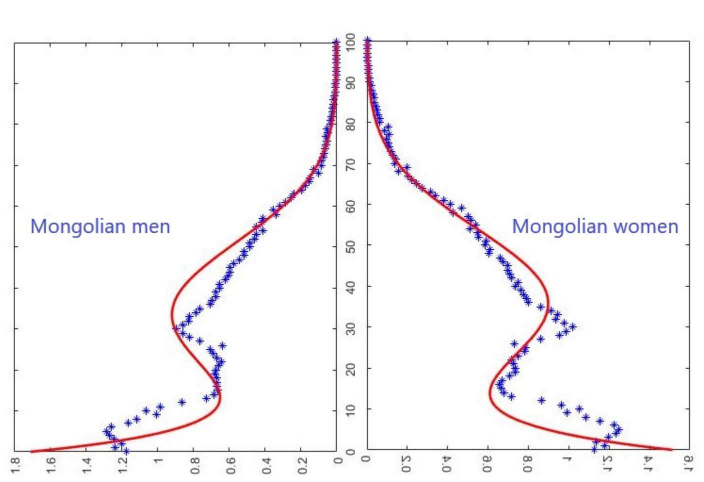

Figure 1. Pyramid population in Mongolia.

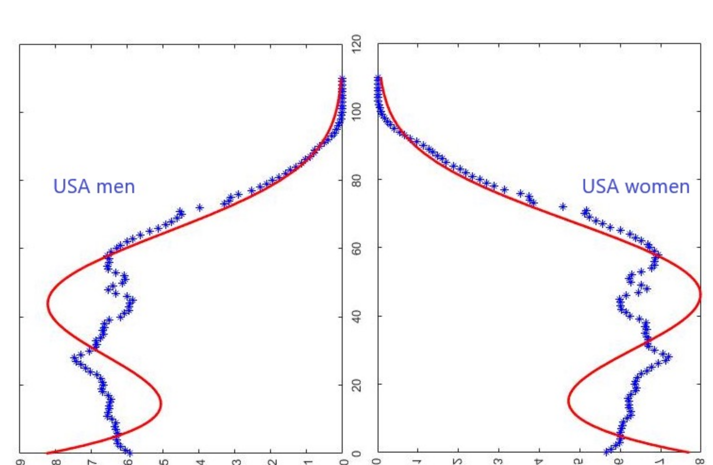

Figure 2. Pyramid population in USA.

Methods

X , Y are age and life span of human, respectively. Let X , Y be random variables with an absolute distribution $functions\ G(x,\tau) = \int_0^x g(t,\tau)dt$, $F(y,\theta) = \int_{\{0\}}^y f(t,\theta)dt$ respectively. Where, $\tau = (\tau_1, \tau_2, \dots, \tau_r) \in \theta_1 \subset R^r$, $\theta = (\theta_1, \theta_2, \dots, \theta_s) \in \Theta_2 \subset R^s$ are unknown vector parameters.
We made the estimation of distribution life span using the sample taken from Y.
In this study, based on the censored data that contain extensive information than our previous study, the distribution of Y is estimated.

The data $Z = \min(X,Y), \delta = \begin{cases} 1, & X \le Y \\ 0, & X > Y \end{cases}$ Is called censored data and its joint probability density function is defined by

$$\tilde{f}(z,k,\tau,\theta) = \left(\left(1 - F(z,\theta)\right)g(z,\tau)\right)^k \left(f(z,\theta)\left(1 - G(z,\tau)\right)\right)^{1-k}, \qquad k = 0,1$$

For censored data $(X_1, Y_1), \dots (X_n, Y_n)$ sample is taken from (X, Y) distributed population corresponding to $(Z_1, \delta_1), \dots, (Z_n, \delta_n)$. The parameters τ, θ based on a sample observations are estimated by the maximum likelihood estimation method

$$L(Z_1, \delta_1, \dots, Z_n, \delta_n, \tau, \theta) = \sum_{i=1}^{n} \ln \tilde{f}(z,,k,\tau,\theta) \to max$$

Pyramid population

We chose $G(x,\tau)$ function to be combined distribution of exponential and normal distributions.

$$g(x,\lambda,\mu,\sigma^2,\alpha) = \alpha\lambda\exp(-\lambda x) + (1-\alpha)\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$
$$x \ge 0, \qquad \lambda > 0, \qquad \sigma^2 > 0, \qquad 0 \le \alpha \le 1$$

The parameters based on a sample observations X are estimated by the maximum likelihood estimation method.
Evaluation of $G(x,\lambda,\mu,\sigma^2,\alpha)$ parameters estimated by maximum likelihood estimation tested with criteria of the modified chi-squared goodness of the fit statistic.

Table 1. Evaluation of the parameter $G(x, \lambda, \mu, \sigma^2, \alpha)$ for some countries

		$\tilde{\lambda}$	$\tilde{\mu}$	$\tilde{\sigma}$	$\tilde{\alpha}$
USA	men	0.050	48.26	18.69	0.36
	women	0.048	50.49	19.22	0.35
Russia	men	0.058	43.50	17.48	0.27
	women	0.053	48.77	18.99	0.25
Japan	men	0.048	53.88	18.82	0.25
	women	0.045	57.37	19.70	0.24
Mongolia	men	0.034	35.98	13.1	0.9
	women	0.032	37.08	13.5	0.9
German	men	0.050	50.29	18.66	0.25
	women	0.048	53.29	19.43	0.24

Estimation of Force of mortality with censored data

We use a third-order B-spline function to construct the logarithm for force of mortality of living time Y.

$$\ln(\lambda(y)) = B(y,\theta,C_k)$$

The number of the knots-k, their locations and B-spline coefficients-θ based on a sample observations are estimated by the maximum likelihood estimation method.

$$L(Z_1,\delta_1,\dots,Z_N,\delta_N,\theta) = \sum_{i=1}^{n_1} \ln\left(\exp\left(-\int_0^{X_i}\lambda(t)dt\right)\cdot g(X_i)\right) + \sum_{i=n_1+1}^{n_1+n_2} \ln\left(\exp\left(B(y,\theta,c_k) - \int_0^{Y_i}\lambda(t)dt\right)(1 - G(Y_i))\right)$$

$$\theta = \{\theta_1,\dots,\theta_p\}, (p = 3 + k); \ C_k = \{c_1,\dots,c_k\}, (a < c_1 < \dots < c_k < b)$$

Evaluation of B-spline parameters estimated by maximum likelihood estimation tested with criteria of the modified chi-squared goodness of the fit statistic. At n-th order B-spline S_k is H_k: $\ln(\lambda(y)) \in S_k$, for $S_k = \{f(y,\theta,c_k): \theta \in R^{\{3+k\}}, C_k \in R^{\{6+k\}}, k \ge 0\}$.
If $\hat{\theta}$, $\hat{C}_k$ denotes the Maximum Likelihood Estimation of the parameter θ, C_k based on the sample, then the modified statistic $\chi^2_{\{N,k\}}(\hat{\theta},\hat{C}_k)$. If the null hypothesis H_k is true, then as the sample size $N \to \infty$, the distribution function of $\chi^2_{\{N,k\}}$ converges to the distribution function of the $\chi^2_{\{N-1\}}$ chi-square distribution with $(N-1)$ -th degree of freedom.
Based on demographic data of the America (2019), Japan (2019), Germany (2019), Russia (2014),and Mongolia(2019) we calculate the estimation and criteria using Matlab.

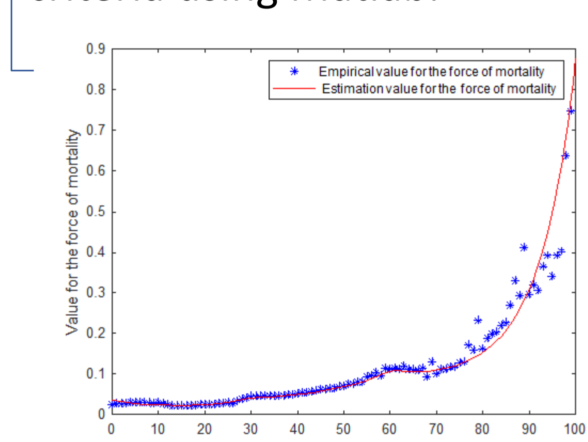

Figure 3. Estimation of Force of mortality for Mongolia

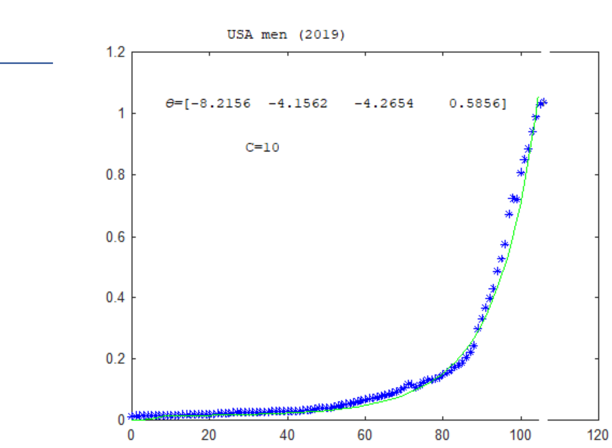

Figure 4. Estimation of Force of mortality for USA

Conclusions

In this study, we constructed the approximate function living time distribution and the force of mortality with censored data.
We chose $G(x,\tau)$ function to be combined distribution of exponential and normal distributions. In the current context, we estimate the logarithm of the force of mortality in the form of the third order B-spline. Its parameters have been estimated by the maximum likelihood estimation. We examine the fit of the obtained distribution with the empirical distribution by using sequential procedure for the modified χ^2 goodness of fit statistics. The results of the study have the advantage of constructing an optimal model with low variance when the amount of data is large. It also drastically reduces the number of knots in the B-spline function, making the design easier.

Contact

[Tserenbat O, Gereltuya.T, Narangoo.G]
[National University of Mongolia, Mongolian University of Science and Technology]
[O_tsbat@yahoo.com, Gereltuya.t@must.edu.mn]

References

1. Tserenbat Oirov, Gereltuya Terbish, Nyamsuren Dorj, "B-spline Estimation for Force of Mortality," Mathematics and Statistics, Vol.9, No.5, pp. 736-743, 2021. DOI: 10.13189/ms.2021.090513
2. D.M.Chibisov, Certain Chi-Square Type Tests for Continuous Distributions,Theory of Probability and Its Applications, Vol. 16, No. 1, pp. 1-22, 1971, https://doi.org/10.1137/1116001
3. James, G., Witten, D., Hastie, T., Tibshirani, R. (2021). Survival Analysis and Censored Data. In: An Introduction to Statistical Learning. Springer Texts in Statistics. Springer, New York, NY. https://doi.org/$10.1007/978-1-0716-1418-1_11$
4. The Human Mortality Database https://www.mortality.org/
5. Mongolian statistical information service https://www.1212.mn/