روش‌های برآورد احتمال مرگ و میر[1]

ابوالفضل طارمی

مقدمه

یکی از روش‌های برآورد احتمالات مرگ و میر که قبلاً مورد توجه بوده اما در حال حاضر به لحاظ محدودیت‌ها و مشکلاتی که داشته کمتر مورد نظر است، روش مشاهدهٔ مستقیم طول عمر یک جمعیت مشخص در طی چندین سال و یافتن احتمال مرگ و میر افراد در سنین مختلف از طریق محاسبهٔ نسبت‌های مرگ و میر در هر سن و هر سال و تعیین جداولی است که برای هر سن احتمال مرگی را مشخص می‌کنند. این جداول، همان جداول مرگ و میر فعلی هستند. معروف‌ترین نمونهٔ این جدول‌ها، جدول مرگ و میر فرانسه است. روش دیگری که امروزه بیشتر مورد توجه واقع شده، نمونه‌گیری از طول عمر افراد در سنین مختلف در یک دورهٔ مشخص و استنباط‌هایی از آن نمونه برای رسیدن به یک برآوردکنندهٔ مناسب جهت احتمال مرگ و میر است. نمونه‌گیری از طول عمر افراد یک جامعه تفاوت عمده‌ای با هر نوع نمونه‌گیری دیگر دارد که به طور مختصر ذکر خواهد شد. اول آن که شرط تصادفی بودن نمونه‌ها در هر نمونه‌گیری روی خود نمونه‌ها اعمال می‌شود در حالی که در نمونه‌گیری طول عمر افراد این شرط فقط در تعیین زمان و دورهٔ نمونه‌گیری دخیل است. بعد از تعیین زمان یا دورهٔ نمونه‌گیری، نمونه‌ها (افراد در دوره) تمام شماری خواهند شد، یعنی طول عمر همهٔ افراد در دوره مطالعه می‌شود. تفاوت دوم این است که در هر نوع نمونه‌گیری، بعد از یافتن نمونه‌ها اطلاعات اندازه‌گیری شدهٔ مشخصی از نمونه‌ها به دست خواهد آمد، در حالی که در

---

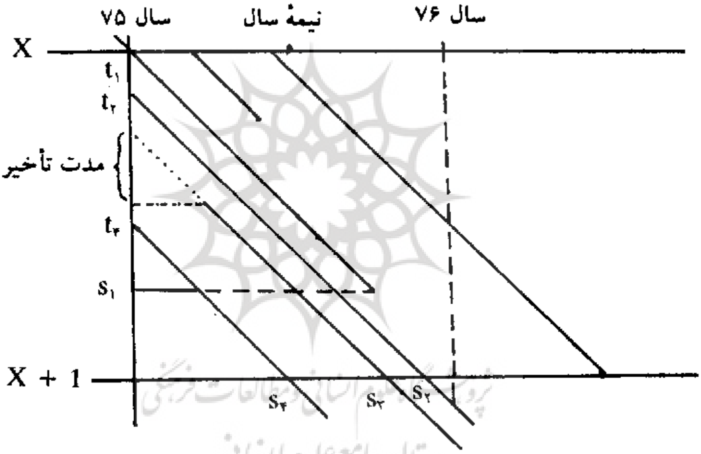
۱. نویسنده از دانشگاه آزاد اسلامی، این عنوان را برای دریافت درجهٔ کارشناسی ارشد در رشتهٔ مدیریت صنعتی برگزیده است.

نمونه‌گیری از طول عمر اندازهٔ متغیر تصادفی طول عمر در دورهٔ مشخص نمونه‌گیری کامل نخواهد بود. برای مثال، بخشی از طول عمر افراد خارج از دورهٔ مشاهدات خواهد بود. یعنی در واقع اطلاعات نمونه‌ای کاملی نداریم.

نمودار لگزیز (بلوک تصادفی)

نمودار لگزیز[1] در واقع به طور شهودی یک بلوک تصادفی است که در آن جریان طول عمر همهٔ افراد در یک سن مشخص (به طور خاص) مشاهده می‌شود. زمان ورود و خروج افراد و یا احیاناً زمان مرگ آنان در بلوک مشهود است.

نمودار ۱. نمودار لگزیز



طول دورهٔ مشاهده را می‌توانیم با توجه به تعداد مشاهدات به دلخواه، یک یا چند سال اختیار کنیم. زمان ورود افراد به دورهٔ مشاهده در سن $x+t_i$ و زمان خروج از دوره یا مرگ در داخل دوره به صورت $x+s_i$ نشان داده می‌شود. طول عمر فرد در دورهٔ مشاهده اصطلاحاً طول عمر آتی فرد در بدو ورود به دوره گفته می‌شود.

تعریف: طول عمر آتی یک فرد در زمان ورود به دورهٔ مشاهده به صورت متغیر تصادفی $T(x+t_i)=s_i-t_i$ در نمودار تعریف می‌شود. مشاهدات ما در نمودار لگزیز در

1. Lexis diagram

واقع مقادیر متغیر تصادفی $T(x+t_i)$ هستند. با این توضیح که افراد در دورهٔ مشاهده به سه دسته تقسیم می‌شوند: دستهٔ اول در داخل دورهٔ مشاهده خواهند مرد؛ دستهٔ دوم مرگشان در خارج از دورهٔ مشاهده اتفاق می‌افتد و ما تنها بخشی از طول عمر آیندهٔ آن‌ها را مشاهده می‌کنیم و بقیهٔ عمرشان بریده شده است؛ و دستهٔ سوم، که با تأخیر تحت مشاهده واقع می‌شوند (به فرض، بیمه شدگانی که بعد از آن‌که مدتی از x سالگی آن‌ها می‌گذرد تحت پوشش بیمه‌ای قرار می‌گیرند). برای این که کلیت مسأله حفظ شود به طور کلی زمان ورود و خروج افراد در دورهٔ مشاهده را در نظر می‌گیریم.

توزیع مشاهدات

برای تعیین توزیع مشاهدات یا در واقع تابع درستنمایی، فرض می‌کنیم که در بلوک تصادفی (نمودار لگزیز) n نفر در یک سن مشخص x سالگی مشاهده شده‌اند. طول عمر آتی آنان را بعد از ورود به بلوک به مدت یک‌سال (طول دورهٔ مشاهده) مشاهده می‌کنیم. در این صورت مشاهداتی که عمرشان در داخل نمودار به مرگ منجر شده است می‌توانند مقادیری برای متغیرهای تصادفی $T_۱,T_۲,...,T_k$ باشند، اندیس این مشاهدات را در مجموعهٔ I مشخص می‌کنیم.

$$T(x+t_i)=s_i-t_i \qquad , \qquad i\in I$$

فرض می‌کنیم که $T_i \sim f_T(t)$. از طرف دیگر $n\text{-}k$ مشاهدهٔ دیگر وجود دارند که طول عمر آن‌ها در نمودار خاتمه نیافته است. مجموعهٔ اندیس‌های این مشاهدات را I^c می‌نامیم. فرض می‌کنیم که $C_۱, C_۲, ..., C_{n-k}$ متغیرهای تصادفی مربوط به مقادیر این مشاهدات باشد.

$$C_i=s_i-t_i \ , \ i\in I^c$$

C_iها متغیرهای تصادفی مربوط به طول عمرهایی است که بخشی از آن‌ها سانسور شده است. فرض می‌کنیم که این متغیرها چگالی احتمال $C_i \sim g(c)$ داشته باشند.

در حالت کلی مشاهدات را به صورت $(Y_۱,\sigma_۱),(Y_۲,\sigma_۲),...,(Y_n,\sigma_n)$ تعریف می‌کنیم که در آن

$$Y_i = min\,(T_i\,, C_i)$$

$$\sigma_i = \begin{cases} ۱ & T_i \le C_i \\ ۰ & T_i > C_i \end{cases}$$

تابع چگالی احتمال توام مشاهدات را با استفاده از تابع توزیع آن‌ها به دست می‌آوریم:

$$H(y) = P(Y \le y) = P(min\,(T_i\,, C_i) \le y) = ۱ - P(T_i \ge y)\,.\,P(C_i \ge y)$$

$$۱ - (۱ - F_{T_i}(y)\,)\,(۱ - G_{C_i}(y)\,) = F_{T_i}(y) + G_{C_i}(y) - F_{T_i}(y)\,G_{C_i}(y)$$

$$h(y) = \frac{\partial H(y)}{\partial y} = f_{T_i}(y) + g_{C_i}(y) - F_{T_i}(y) g_{C_i}(y) - f_{T_i}(y)\,G_{C_i}(y)$$

اگر داشته باشیم $\sigma_i = ۱$، آن‌گاه داریم:

$$T_i \le C_i \quad \Rightarrow \quad Y_i = T_i \quad \Rightarrow \quad ۱ - F_{T_i}(y) = ۰$$

$$\Rightarrow h(y) = f_{T_i}(y)\,(\,۱ - G_{C_i}(y))$$

اگر داشته باشیم $\sigma_i = ۰$، آن‌گاه داریم:

$$T_i > C_i \quad \Rightarrow \quad Y_i = C_i \quad \Rightarrow \quad ۱ - G_{C_i}(y) = ۰$$

$$\Rightarrow h(y) = g_{C_i}(y)\,(\,۱ - F_{T_i}(y))$$

تابع چگالی توام مشاهدات به صورت حاصل‌ضرب توابع چگالی آن‌ها به دست می‌آید:

$$L(y\,, \Theta) = \prod_{i=۱}^{n} h_\Theta(y_i) = \prod_{i \in I} f_{T_i}(y_i)\,(\,۱ - G_{C_i}(y_i)\,) \prod_{i \in I^c} g_{C_i}(y_i)\,(\,۱ - F_{T_i}(y_i)\,)$$

$$= \prod_{i \in I} f_{T_i}(y_i) \prod_{i \in I^c} (\,۱ - F_{T_i}(y_i)\,) \underline{\prod_{i \in I^c} g_{c_i}(y_i) \prod_{i \in I} (\,۱ - G_{C_i}(y_i)\,)}$$

زمانی که توزیع متغیرهای تصادفی $C_1, C_2, ..., C_n$ به پارامترهای مجهول طول عمر آینده وابسته نباشد مقدار مشخص شده در معادلهٔ فوق مقدار ثابتی خواهد داشت و در تعیین برآوردکنندهٔ مربوط به پارامتر طول عمر نقشی نخواهد داشت و می‌توانیم آن را حذف کنیم. بنابر اصل درستنمایی داریم:

$$L(\underset{\sim}{y}, \theta) \propto \prod_{i \in I} f_{T_i}(y_i) \prod_{i \in I^c} (1 - F_{T_i}(y_i))$$

می‌دانیم تابع چگالی احتمال متغیر تصادفی $T(x)$ عبارت است از $g(t)$ که از رابطهٔ $\mu_{x+t} = \frac{g(t)}{1 - G(t)}$ می‌توانیم آن را به دست آوریم:

$$g(t) = \mu_{x+t} \cdot {}_tp_x$$

بنابر این رابطهٔ تابع چگالی احتمال متغیر تصادفی $T(x+t_i) = s_i - t_i = y_i$ برابر است با

$$g_{T_i}(y_i) = \mu_{x+t_i+y_i}({}_{y_i}p_{x+t_i}) = \mu_{x+s_i}({}_{s_i-t_i}p_{x+t_i})$$

پارامتر θ در تابع درستنمایی مورد بحث μ_{x+s_i}، نیروی مرگ و میر است:

$$L(\theta, \underset{\sim}{y}) = \prod_{i \in I} \mu_{x+t_i+y_i} \cdot {}_{y_i}p_{x+t_i} \prod_{i \in I^c} {}_{y_i}p_{x+t_i}$$

با فرض $\theta = \mu_{x+t_i+y_i}$ و $y_i = s_i - t_i$ داریم:

$$L(\theta, \underset{\sim}{y}) = \prod_{i \in I} \mu_{x+t_i+y_i} \cdot {}_{y_i}p_{x+t_i} \prod_{i \in I^c} {}_{s_i-t_i}p_{x+t_i}$$

با پذیرفتن فرض ثابت بودن نیروی مرگ و میر در طول یک سال (فرض دوم احتمال مرگ و میر) تابع درستنمایی شکل ساده‌تری به خود می‌گیرد:

$$\mu_{x+s_i} = \mu_{x+0/5} \qquad 0 < s_i < 1$$

$$L(\mu_{x+0/5}, \underset{\sim}{y}) = \prod_{i \in I} \mu_{x+0/5} \prod_{i=1}^{n} {}_{y_i}p_{x+t_i}$$

اگر تعداد فوت شدگان دورهٔ مشاهده با نماد d_x در نظر گرفته شود:

$$L(\mu_{x+۰/۵}, \underset{\sim}{y}) = (\mu_{x+۰/۵})^{d_x} \exp\left(-\mu_{x+۰/۵} \sum_{i=۱}^{n} y_i\right)$$

روش‌های کلاسیک برای برآورد احتمال مرگ

روش اول (روش *MLE*): در این روش از تابع درستنمایی به دست آمده از رابطهٔ اخیر استفاده می‌کنیم. با این توضیح که از روش *MLE* تنها در صورتی می‌توان استفاده کرد که نیروی مرگ و میر در طول یک سال مقدار ثابتی باشد. از طریق بیشینه کردن تابع درستنمایی برآورد کننده‌ای برای پارامتر مجهول $\mu_{x+\frac{۱}{۲}}$ به دست می‌آوریم:

$$l(\mu_{x+۰/۵}, \underset{\sim}{y}) = d_x \ln(\mu_{x+۰/۵}) - \left(\mu_{x+۰/۵} \sum_{i=۱}^{n} y_i\right)$$

$$\frac{\partial l(\mu_{x+۰/۵}, \underset{\sim}{y})}{\partial(\mu_{x+۰/۵})} = ۰ \quad \rightarrow \quad \frac{d_x}{\mu_{x+۰/۵}} = \sum_{i=۱}^{n} y_i$$

$$\hat{\mu}_{x+۰/۵} = \frac{d_x}{\sum_{i=۱}^{n} y_i}$$

$$q_x = ۱ - p_x = ۱ - \exp\{-\mu_{x+۰/۵}\} \rightarrow \hat{q}_x = ۱ - \exp\{-\hat{\mu}_{x+۰/۵}\}$$

$$\hat{q}_x = ۱ - \exp\left\{-\frac{d_x}{\sum_{i=۱}^{n} y_i}\right\}$$

روش دوم: با استفاده از امید ریاضی تعداد مرگ‌ها در نمودار لگزیز و استفاده از فرض سوم احتمال مرگ و میر (فرض بالدوچی) که طبق آن فردی که هم‌اکنون یک‌سال از عمر خود را می‌گذراند (قبلاً ۲۵ سال و ۳ ماه سن دارد) احتمال مرگش در مدت زمان باقی‌مانده تا آخر سال (${}_{۱-\frac{۳}{۱۲}}q_{۲۵+\frac{۳}{۱۲}}$) برابر حاصل‌ضرب این مدت در احتمال مرگ در طی یک‌سال مورد بحث است ($(۱-\frac{۳}{۱۲})\,.\,{}_{۱}q_{۲۵}$)، برآورد کننده‌ای برای احتمال مرگ و میر افراد x ساله به دست می‌آوریم.

تعریف: D_x امید ریاضی تعداد فوت‌های مربوط به سن X سال در طی یک‌سال و d_x تعداد فوت‌های مشاهده شده در نمودار لگزیز است:

$$D_x = \sum_{i=1}^{n} {}_{1-t_i}q_{x+t_i} - \sum_{i \notin I} {}_{1-s_i}q_{x+s_i}$$

رابطهٔ فوق بیانگر این مطلب است که مجموع احتمالات مربوط به فوت همهٔ افراد تا پایان سال منهای مجموع احتمالات فوت افرادی که تا پایان سال در نمودار عمر می‌کنند، برابر امید ریاضی تعداد فوت‌هاست. با پذیرفتن فرض بالدوچی:

$$\begin{cases} {}_{1-t_i}q_{x+t_i} = (1-t_i)q_x \\ {}_{1-s_i}q_{x+s_i} = (1-s_i)q_x \end{cases}$$

$$D_x = \sum_{i=1}^{n}(1-t_i)q_x - \sum_{i \notin I}(1-s_i)q_x = \left(\sum_{i=1}^{n}(s_i-t_i) + \sum_{i \in I}(1-s_i)\right).q_x$$

$$\hat{q}_x = \frac{\hat{D}_x}{\sum_{i=1}^{n}(s_i-t_i) + \sum_{i \in I}(1-s_i)}$$

اشکال اساسی این برآورد آن است که ممکن است گاهی مقدار احتمال بیشتر از یک باشد و آن زمانی است که تعداد مرگ‌ها به نسبت طول عمرهای داخل نمودار بیشتر باشد. البته باید توجه داشت که مقادیر صورت کسر صحیح و مخرج مقدار حقیقی خواهد بود.

روش پیشنهادی برای تعیین احتمالات مرگ و میر (روش کلاسیک)

در این روش با استفاده از شیوهٔ نمونه‌گیری خاصی، یک سری اطلاعات نمونه‌ای از جامعه به دست می‌آوریم. با تعریف متغیر تصادفی مناسبی در نمونه‌های به دست آمده توزیع آن را پیدا می‌کنیم. سپس با ارائه یک برآورد کنندهٔ مناسب از روی مشاهدات و مقایسهٔ آن با برآوردکننده‌های معمول، شرایط و خصوصیات برتری آن را بررسی می‌کنیم.

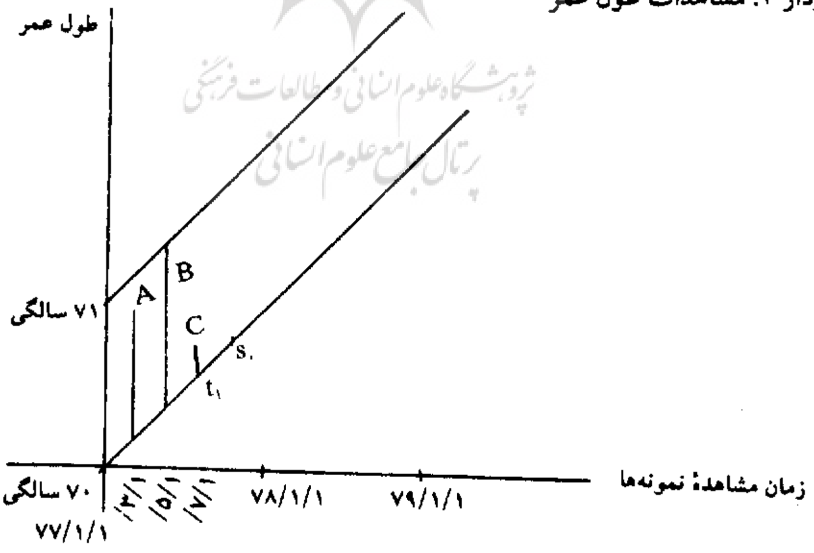
روش نمونه‌گیری

در یک جامعهٔ آماری مشخصی از افراد، توزیع مشترک طول عمر تعدادی از آن‌ها را

در نظر می‌گیریم. با انتخاب یک فاصلهٔ زمانی مشخص و دلخواه در یک مقطع زمانی که به طور تصادفی انتخاب می‌شود به‌بررسی طول‌عمر افراد از بدو ورود به‌دورهٔ مشاهدات تا پایان دوره می‌پردازیم.

یک سن مشخص (x) را در نظر می‌گیریم (در یک دورهٔ مشاهده، معمولاً افراد با سنین مختلف وجود دارند که تنها یک سن را مورد توجه قرار می‌دهیم؛ برای بقیهٔ سنین هم می‌توان از همین نمونه استفاده کرد). برای مثال، دورهٔ زمانی ۱۳۷۷/۱/۱ تا ۱۳۷۷/۷/۱ را می‌توان مقطع زمانی برای دورهٔ نمونه‌گیری در نظر گرفت. در این دوره تعداد معین و دلخواهی از همهٔ افرادی را که به نحوی به سن xسالگی می‌رسند در نظر می‌گیریم و تا پایان زمان رسیدن به سن $x+1$سالگی تعقیب می‌کنیم. دورهٔ مشاهدات از زمان شروع نمونه‌گیری تا پایان مشاهده حداقل باید یک سال باشد. مشاهدات طول عمر افراد را به صورت نمودار شمارهٔ ۲ در محور مختصات می‌توانیم نمایش دهیم. محور افقی در این نمودار نشان دهندهٔ زمان مشاهدهٔ نمونه‌ها بر حسب ترتیب ورود به دورهٔ نمونه‌گیری و محور عمودی طول عمر افراد از سن x تا $x+1$سالگی را نشان می‌دهد. زمانی را که سن شخص iام به xسالگی می‌رسد با متغیر t_i و زمانی را که شخص به سن $x+1$سالگی می‌رسد و یا احیاناً می‌میرد با s_i نشان می‌دهیم.

نمودار ۲. مشاهدات طول عمر



فرد «الف»، دو ماه بعد از شروع دورهٔ نمونه‌گیری وارد دوره می‌شود و ده ماه بعد از ۷۰ سالگی عمر می‌کند.

فرد «ب»، چهار ماه بعد از شروع دوره نمونه‌گیری وارد دورهٔ مشاهدات می‌شود و یک سال کامل بعد از ۷۰ سالگی عمر می‌کند.

فرد «پ»، شش ماه بعد از شروع دوره نمونه‌گیری وارد دوره می‌شود و دو ماه بعد از ۷۰ سالگی می‌میرد.

متغیر تصادفی ما در جریان نمونه‌گیری عبارت است از $Y_i = s_i - t_i$ ، طول عمر افراد در بدو ورود به xسالگی تا زمان $x+1$سالگی که گاهی ممکن است فرد کمتر از یک سال عمر کند. به وضوح متغیر تصادفی، یا مقدار ثابت یک را به خود می‌گیرد یا متغیر تصادفی مقداری کمتر از یک است که در واقع طول عمر آتی فرد در این حالت محسوب می‌شود.

$$Y_i = \begin{cases} 1 & \text{اگر شخص در فاصلهٔ } (x\,,\,x+1) \text{ زنده باشد.} \\ T(x_i) = s_i - t_i & \text{اگر شخص } x \text{ساله کمتر از یک سال دیگر عمر کند.} \end{cases}$$

چنانچه n فرد در دورهٔ مشاهده، نمونه‌گیری شده باشند و بخواهیم تابع درستنمایی مربوط را به دست آوریم. به طور مشابه با توجه به روش نمودار لگزیز، تابع درستنمایی به شکل زیر خواهد بود:

$$L(\underset{\sim}{y}\,,\Theta) = \prod_{i \in I} f_T\,(y_i) \prod_{i \in I^c} (1 - F_T\,(y_i))\;\;\Phi(y_i)$$

برآورد کنندهٔ روش درستنمایی برای این مشاهدات، همان برآورد کننده روش MLE مربوط به نمونه‌گیری به روش لگزیز خواهد بود:

$$\hat{\mu}_{x+\frac{1}{2}} = \frac{d_x}{\sum_{i=1}^{n} y_i}$$

$$\hat{q}_x = 1 - exp\,\{-\,\hat{\mu}_{x+\frac{1}{2}}\}$$

برآورد احتمال مرگ و میر افراد در یک سن خاص

۱. روش نسبتی

روش متعارف برای یافتن احتمال مرگ و میر این است که از نسبت تعداد افراد فوت شده

در طی یک‌سال (از x تا $x+1$ سالگی) به کل افراد ابتدای سال استفاده کنیم یعنی

$$ {}_{۱}\hat{q}_x = \frac{{}_{۱}d_x}{n} $$

آن‌چنان که در جداول مرگ و میر فعلی و جدول مرگ و میر فرانسه از این روش استفاده شده است.

۲. استفاده از نسبت طول عمر افراد در طی یک‌سال به عنوان برآوردگر احتمال بقا

مشاهدات y_i را قبلاً به صورت عمر آتی فرد i ام بعد از سن x سالگی تعریف کردیم. با توجه به نمودار شمارهٔ ۲ نسبت طول عمرهای افراد در طی یک سال به زمان‌های مربوط به عمر افراد در طول یک‌سال (در یک سن خاص) به صورت

$$ P = \frac{\sum_{i=۱}^{n} y_i}{n} $$

است.

اگر همهٔ n فرد عمر یک ساله کامل داشته باشند مجموعهٔ همهٔ طول عمرهای یک ساله برابر با n سال خواهد بود. در صورت کسر، حالات مساعد (طول عمر) در نظر گرفته شده است. از این نسبت می‌توان به منزلهٔ برآورد کننده‌ای برای احتمال بقا در آن سن استفاده کرد.

با یک مثال سادهٔ دو روش فوق را با هم مقایسه می‌کنیم.

اگر در یک نمونهٔ ده نفری که x ساله هستند در سال آخر عمرشان مقادیر بردار طول عمر آن‌ها (بر حسب سال) به صورت زیر باشد:

$$ \underline{y} = (۱, ۱, ۰/۷۵, ۱, ۱, ۱, ۰/۵, ۱, ۰/۷۵, ۱) $$

برای مثال، اگر مقدار y برابر ۰/۵ باشد به این معناست که فرد در سن x سالگی دارای طول عمر آتی شش ماهه بوده است).

با استفاده از روش نسبت تعداد مرگ‌ها به کل افراد ابتدای دوره داریم:

$$ {}_{۱}q_x = \frac{d_x}{n} = \frac{۳}{۱۰} = ۰/۳ \qquad , \qquad p_x = ۰/۷ $$

چنانچه از نسبت طول عمر بخواهیم به منزلهٔ احتمال بقا استفاده کنیم مقدار آن به صورت زیر است:

$$ {}_{1}p_x = p = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{9}{10} = 0/9 \qquad , \qquad {}_{1}q_x = 0/1 $$

بین مقادیر دو نسبت، تفاوت چشم‌گیری وجود دارد و ملاحظه می‌شود که روش دوم روش دقیق‌تری است.

روش بیزی برای برآورد احتمال مرگ و میر

روش کار بیزی به این صورت است که اگر بخواهیم اطلاعاتی در مورد مقدار یک پارامتر مجهول از یک جامعه به دست آوریم، آن پارامتر را به منزلهٔ یک کمیت تصادفی در نظر می‌گیریم و با توجه به تجربیات گذشته و اطلاعات پیشین، توزیع اولیه‌ای برای آن در نظر می‌گیریم. سپس به کمک اطلاعات حاصل از یک نمونهٔ تصادفی از جامعه‌ای که توزیع آن به پارامتر فوق وابسته است توزیع دیگری را از آن کمیت به نام توزیع پسین به دست می‌آوریم. سرانجام به کمک این توزیع پسین، برآورد کننده‌ای از آن پارامتر به دست می‌آوریم.

در این بخش می‌خواهیم احتمال مرگ افراد گروه سنی ۷۰ تا ۷۴ سالهٔ شهر تهران طی یک‌سال آینده را به‌دست‌آوریم. متغیر تصادفی طول‌عمر آیندهٔ افراد را قبلاً $T(x)=y$ در نظر گرفتیم که چگالی احتمال آن بر حسب نیروی مرگ به‌صورت $g(y_i \mid \mu_{x+y_i}) = \mu_{x+y_i}\, {}_{y_i}p_x$ است.

با لحاظ کردن فرض ثابت بودن نیروی مرگ در طول یک سال توزیع توام مشاهدات، y_iها، به صورت زیر است:

$$ L(\mu_{x+\frac{1}{2}}, \underline{y}) = (\mu_{x+\frac{1}{2}})^{d_x} exp\left\{ -\mu_{x+\frac{1}{2}} \sum_{i=1}^{N} y_i \right\} $$

تابع درستنمایی فوق در حالت کلی بدون در نظر گرفتن توزیع y_iهاست. هم‌چنان‌که چگالی احتمال $g(y_i)$ هم به‌صورت یک چگالی احتمال به‌فرم کلی در نظر گرفته شده است.

تغییرات متغیر تصادفی طول عمر آتی افراد ناشی از دو عامل نیروی مرگ و احتمال بقای افراد است. از این رو، صرف نظر از نوع چگالی احتمال y_iها از فرم کلی استفاده شده است.

متغیر تصادفی Y همان‌طور که قبلاً نیز گفته شد برای فرد iام به صورت

$$Y_i = \begin{cases} ۱ & \text{اگر شخص در فاصلهٔ } (x , x+1) \text{ زنده باشد.} \\ y_i = s_i - t_i & \text{اگر شخص } x \text{ساله کمتر از یک سال دیگر عمر کند.} \end{cases}$$

تعریف شده است.

ملاحظه می‌شود که y یک توزیع آمیخته دارد. بخشی از جرم احتمال در نقطهٔ یک متمرکز شده است که اغلب این نقطه احتمال بیشتری به خود اختصاص می‌دهد (به علت این که اغلب افراد در گروه‌های سنی مختلف، یک سال دورهٔ مشاهدات را سپری می‌کنند) و بقیهٔ احتمال روی مقادیر کمتر از یک در یک قالب مشخص احتمالی توزیع شده‌اند که ما به این قسمت بیشتر توجه می‌کنیم.

توزیع بتا با توجه به خواص انعطاف پذیری آن در فاصلهٔ تغییرات (۰،۱) مناسب‌ترین توزیع برای برازش بر این نوع متغیر تصادفی است.

برای y_iهای مربوط به افراد فوت شده در سنین ۷۰ تا ۷۴ سالگی، آزمون نیکویی برازش توزیع بتا انجام گرفته است. نتایج به دست آمده بیانگر این واقعیت است که توزیع بتا با درجه بالایی از اطمینان قابلیت برازش بر چنین داده‌هایی را داراست.

توزیع توأم مشاهدات (y_iها) را اگر بخواهیم با توجه به توزیع y_iها به دست آوریم خواهیم داشت:

$$L(\alpha,\beta,y) = \left(\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\right)^N \prod_{i=۱}^{N}\left[y_i^{\alpha-۱}(۱-y_i)^{\beta-۱}\right] \;,\; ۰\le y_i\le ۱ \;,\; \alpha,\beta>۰$$

پرسشی که این‌جا ممکن است پیش بیاید این است که این رابطه چگونه با رابطهٔ زیر یعنی رابطهٔ کلی (بدون در نظر گرفتن فرض ثابت بودن نیروی مرگ) برابر خواهد بود؟

$$L(\mu_{x+t_i},y) = \prod_{i\in I} \mu_{x+t_i+y_i}\, exp\left[-\sum_{i=۱}^{N}\int_{۰}^{y_i}\mu_{x+t_i+s}\,ds\right]\psi(y)$$

نکتهٔ اساسی که می‌تواند این دو تابع را به هم ربط دهد این است که نیروی مرگ یک رابطهٔ از پارامترهای β,α است، به نحوی که هر دو تابع برابر شوند. به دلیل این که به دست آوردن رابطه بین نیروی مرگ و β,α قدری دشوار و خارج از حوصلهٔ این بحث است از آن صرف نظر می‌کنیم.

به طور خلاصه با پذیرفتن ثابت بودن نیروی مرگ در طی یک سال، توزیع مشاهدات، y_iها، (با فرض داشتن توزیع بتا) به صورت زیر است:

$$L(\mu_{x+\cdot/\text{۵}}, \underset{-}{y}) \propto (\mu_{x+\cdot/\text{۵}})^{d_x} exp\left[-\mu_{x+\cdot/\text{۵}} \sum_{i=\text{۱}}^{N} y_i\right]$$

برای برآورد کردن احتمال مرگ در سنین ۷۰ تا ۷۴ سالگی از روش بیزی، ابتدا برآورد کننده‌ای برای $\mu_{x+\frac{1}{2}}$ به دست می‌آوریم و از روی آن مقدار احتمال مرگ را در سنین ۷۰ تا ۷۴ سالگی[۱] به دست می‌آوریم.

انتخاب توزیع گاما به‌عنوان توزیع پیشین $\mu_{x+\frac{1}{2}}$ در سنین ۷۰ تا ۷۴ سالگی

توزیع پیشین برای یک متغیر تصادفی باید حاصل[۲] تجربیات و ذهنیات اولیه در مورد آن متغیر باشد. اگر برای هر سن نیروی مرگ ($\mu_{x+\frac{1}{2}}$) را یک متغیر تصادفی θ در نظر بگیریم انتظار داریم شرایط زیر را داشته باشد:

۱. θ مقادیر مثبت اختیار کند $0<\theta<\infty$

۲. حد چگالی احتمال θ در نقاط صفر و بی‌نهایت صفر است.

۳. چگالی احتمال متغیر تصادفی θ تک مدی باشد یعنی در هر سن θ حول یک نقطه متمرکز باشد.

۴. دامنهٔ تغییرات متغیر تصادفی θ در هر سن برای نمونه‌های مختلف، کم باشد.

۵. با افزایش سن، θ به طور طبیعی سیر صعودی داشته باشد.

۶. در سنین ۱۵ تا ۵۰ سالگی که سنین کم بحران نامیده می‌شود مقدار θ نیروی مرگ بین صفر و یک و نزدیک به صفر باشد.

در مورد مقادیر θ برای سنین مورد بحث (۷۰ تا ۷۴ سالگی) از تجربیات حاصل از جداول مرگ و میر فعلی هم لازم است استفاده کنیم.

تلفیقی از اطلاعات فوق می‌تواند ما را در دست‌یابی به یک توزیع پیشین تجربی مناسب یاری کند. نمود عینی این اطلاعات را می‌توانیم به صورت نمودار برای سنین مورد مطالعه ترسیم کنیم.

با توجه به شرایط فوق تنها چگالی که در شرایط یاد شده صدق می‌کند تابع چگالی احتمال گاماست. با این فرض می‌توانیم توزیع پسین و به دنبال آن برآورد کننده‌ای برای

نیروی مرگ و میر به دست آوریم:

$$u(\Theta \mid \alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \Theta^{\alpha-1} e^{-\frac{\Theta}{\beta}} \qquad \Theta>0 \quad , \quad \alpha,\beta>0$$

$$u(\Theta \mid \underset{\sim}{y}) = \frac{\Theta^{d_x+\alpha-1} e^{-\Theta(\sum_{i=1}^{N} y_i + \frac{1}{\beta})}}{\int_0^{\infty} \Theta^{d_x+\alpha-1} e^{-\Theta(\sum_{i=1}^{N} y_i + \frac{1}{\beta})} d\Theta}$$

$$u(\Theta \mid \underset{\sim}{y}, \alpha,\beta) \sim Gamma\left(d_{x+\alpha}, \frac{1}{\sum_{i=1}^{N} y_i + \frac{1}{\beta}}\right)$$

$$\hat{\Theta} = E\left[\Theta \mid \underset{\sim}{y}\right] = \frac{d_{x+\alpha}}{\sum_{i=1}^{N} y_i + \frac{1}{\beta}} \qquad (۱.۱)$$

برای محاسبهٔ مقدار عددی برآورد نیروی مرگ در سنین ۷۰ تا ۷۴ سالگی لازم است پارامترهای α,β در توزیع پیشین به روش‌های ممکن برآورد شود.

برای برآورد α,βی مربوط بـه تـوزیع پیشین گـاما از روش بـیزی تـجربی اسـتفاده می‌کنیم. با استفاده از توزیع طول عمر آتی افراد $g(y \mid \Theta) = \Theta e^{-\Theta y}$ چگالی احتمال کناری مشاهدات را به صورت زیر به دست می‌آوریم: $(\Theta = \mu_{x+0/5})$

$$u(\Theta \mid \alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \Theta^{\alpha-1} e^{-\frac{\Theta}{\beta}} \qquad \Theta>0 \quad , \quad \alpha,\beta>0$$

$$g(y,\Theta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \Theta^{\alpha} e^{-\Theta(y+\frac{1}{\beta})}$$

$$g(y) = \frac{\alpha\beta}{(\beta y+1)^{\alpha+1}} \qquad 0 \le y \le 1 \quad , \quad \alpha,\beta>0$$

اگر یک نمونهٔ تصادفی nتایی از این توزیع داشته باشیم از روش MLE برآورد α,β به

صورت زیر قابل محاسبه است:

$$L(\alpha,\beta,\underset{\sim}{y}) = \prod_{i=1}^{n} \frac{\alpha\beta}{(\beta y_i + 1)^{\alpha+1}}$$

$$l\,(\alpha,\beta,\underset{\sim}{y}) = \sum_{i=1}^{n} \left(ln\alpha + ln\beta - (\alpha+1)\,ln(\beta y_i + 1) \right)$$

$$\frac{\partial l}{\partial \alpha} = 0 \qquad \Rightarrow \ \hat{\alpha} = \frac{n}{\sum_{i=1}^{n} ln\,(\beta y_i + 1)}$$

$$\frac{\partial l}{\partial \beta} = 0 \qquad \Rightarrow \ \hat{\beta} = \frac{n}{\sum_{i=1}^{n} \frac{(\alpha+1) y_i}{\beta y_i + 1}}$$

مقادیر $\hat{\beta}$, $\hat{\alpha}$ در دستگاه معادلات به دست آمده به صورت بسته قابل محاسبه نیست. لذا از روش‌های آنالیز عددی برآورد آن‌ها را به دست می‌آوریم.

بعد از یافتن جواب معادلات در رابطهٔ (۱.۱) به جای α,β مقادیر برآورد شدهٔ آن را قرار می‌دهیم و برای سنین مورد مطالعه (۷۰ تا ۷۴ سالگی) برآورد احتمال مرگ و میر را محاسبه می‌کنیم. برای این که بتوانیم برای برآوردهای احتمال مرگ و میر و نیروی مرگ، میزان دقت و فاصلهٔ اطمینان به دست آوریم، لازم است نمونه‌گیری به عمل آمده از طول عمر افراد به تعداد قابل قبولی تکرار شود و از هر نمونه نیروی مرگ و احتمال مرگ به دست آید، سپس متوسط مقادیر به دست آمده و هم‌چنین انحراف معیار آن‌ها را به دست آوریم.

سن	متوسط برآورد نیروی مرگ	واریانس	فاصله اطمینان		متوسط برآورد احتمال مرگ	فاصله اطمینان	
			کران پایین	کران بالا		کران پایین	کران بالا
۷۰	۰/۰۴۳۳۰	۲/۳۳۵۶۴۸e(-۰/۰۰۹)	۰/۰۴۳۲۱	۰/۰۴۳۳۶	۰/۰۴۲۳۸	۰/۰۴۲۲۹	۰/۰۴۲۴۴
۷۱	۰/۰۴۷۸۸	۲/۲۹۳۲۹۱e(-۰/۰۰۹)	۰/۰۴۷۷۹	۰/۰۴۷۹۶	۰/۰۴۶۷۵	۰/۰۴۶۶۷	۰/۰۴۶۸۳
۷۲	۰/۰۵۵۸۴	۱/۰۶۸۸۸۸e(-۰/۰۰۹)	۰/۰۵۵۷۸	۰/۰۵۵۸۸	۰/۰۵۴۳۱	۰/۰۵۴۲۵	۰/۰۵۴۳۶
۷۳	۰/۰۶۶۵۱۲	۱/۲۷۲۷۸e(-۰/۰۰۹)	۰/۰۶۶۴۴	۰/۰۶۶۵۷	۰/۰۶۴۳۴	۰/۰۶۴۲۹	۰/۰۶۴۴۱
۷۴	۰/۰۷۹۳۳	۴/۱۰۴۱۳۸e(-۰/۰۰۹)	۰/۰۷۹۲۳	۰/۰۷۹۴۵	۰/۰۷۶۲۷	۰/۰۷۶۱۷	۰/۰۷۶۳۸