# ESTIMATING COUPLED LIVES MORTALITY

Doris Dregger

July 15, 2003

# CHAPTER I

# INTRODUCTION

Retirement security systems are designed to provide income in retirement. The majority of workers in the United States retire before age 65. Since they lose their income after retirement they will be financially insecure unless they have sufficient savings or other sources of retirement income. The reduction in income upon retirement can result in a reduced standard of living. For example, in 2000 the median income for households with someone age 65 and over in the United States was 45 percent less than the median for all households in the United States (Rejda, 2003).

For married couples, if one spouse earns significantly more than the other, this person has to take care of his or her own income after retirement. In addition, he or she must consider their spouse's income if their spouse outlives him or her. Apart from living costs, the surviving spouse may have additional expenses such as funeral expenses, uninsured medical bills, estate settlement costs, and federal estate taxes for larger estates.

Economic security for retired workers and survivors of deceased workers in the United States is, in the majority of cases, provided by Social Security, private pensions, and individual savings. This is called *the three pillars of economic-security protection* (Allen et al., 2003).

1.1 Social Security System

The first pillar is the Social Security System (Old Age Security and Disability Income, or OASDI). It is a government-sponsored public retirement program funded by general tax revenue, payroll tax revenue, or both. The program was enacted into law as a result of the Social Security Act of 1935. More than 90 percent of all workers are working in occupations covered by OASDI (Rejda, 2003). The OASDI program pays monthly retirement and disability benefits to eligible beneficiaries and it pays survivor benefits to eligible surviving family members.

1.2 Pension Plans

In the United States, employer-provided pension plans make up the second layer of protection against financial insecurity. Millions of workers participate in private retirement plans. Federal legislation and the Internal Revenue Code have had a great influence on the design of these plans. The Employee Retirement Income Security Act of 1974 (ERISA) established guidelines that affect the tax, investment, and accounting aspects of employer-provided retirement plans. The Taxpayer Relief Act of 1997 and the Economic Growth and Tax Relief Reconciliation Act of 2001 increased the tax advantages of private retirement plans for employers and employees (Rejda, 2003). The Internal Revenue Service (IRS) issues new regulations that affect the design of private retirement plans. A *qualified retirement plan* is defined as a plan that meets the requirements established by the IRS and receives favorable income tax treatment. The employer's contributions (within limits) are tax deductible by the employer as a business expense. These contributions are not considered taxable income to the employees; the investment earnings on plan assets are not subject to federal income tax until paid in the form of benefits (Allen at al., 2003, Rejda, 2003, and Hallman, 2003). Qualified pension plans together with Social Security benefits will generally provide for 50 to 60 percent of the worker's gross earnings before retirement. A qualified plan must benefit all workers regardless of their income. A plan must satisfy certain minimum coverage requirements to be a qualified plan. We will not discuss these requirements in detail. However, since discrimination in favor of highly compensated employees must be avoided, a qualified plan must satisfy one of the following tests, which are described in (Rejda, 2003). (1) Ratio percentage test: If a plan covers a special percentage, $p$, of the highly compensated employees, it must also cover at least 70 percent of $p$ of the non-highly compensated employees. (2) Average benefits test: Two requirements must be fulfilled: (a) The plan must not discriminate in favor of highly compensated employees, and (b) the average benefit for non-highly compensated employees must be at least 70 percent of the average benefit for the highly compensated employees.

If an employee is at least 21 and has one year of service he must be allowed to participate in a qualified retirement plan. To remain qualified, a pension plan cannot force anyone to retire at some mandatory retirement age.

### 1.2.1 Classification of Pension Plans

There are two types of pension plans: *defined benefit plans* and *defined contribution plans*. In a defined contribution plan the contribution rate is defined, but the actual retirement benefit varies depending on the worker's age of entry into the plan, contribution rate, investment rate, and the age of normal retirement. The *normal retirement age* is the age that a worker receives a full, unreduced benefit, when he retires (Rejda, 2003). A defined benefit plan defines the monthly retirement benefit but the contribution varies depending on the amount needed to fund the desired benefit. The employer is expected to have sufficient funds to provide the benefits. The benefits typically depend on both earnings and years of participation in the plan.

In the past, defined benefit plans were more popular than defined contribution plans. With the passage of the Employment Retirement Income Security Act (ERISA) in 1974, defined contribution plans have grown in relative importance.

### 1.2.2 QPSA and QJSA

The Retirement Equity Act (REA) of 1984 requires that an employer provides preretirement death benefits in the form of an annuity to the surviving spouse of a deceased vested participant (called a *qualified preretirement survivor annuity* or QPSA) for married participants. Only spouses of those employees who die before retirement receive this benefit. The payment of the benefit must begin no later than the day on which the deceased would have reached the early retirement age. The *early retirement age* is the earliest age a worker can retire to receive a retirment benefit. In defined benefit plans it is assumed that the participant has terminated employment (instead of dying), survived to the earliest retirement age, retired with an immediate QJSA (see next paragraph) at the earliest retirement age, and dies one day after. In defined contribution plans, the benefit must be an annuity for the surviving spouse which is actuarially equivalent to at least 50 percent of the participant's vested account balance on the day of death.

In addition, REA requires that an employer must provide a *qualified joint and survivor annuity* (QJSA) for a married participant. This annuity pays over the lifetime of the participant and, when the participant dies, continues payments to the surviving spouse. The benefit for the surviving spouse is at least 50 percent and at most 100 percent of the payments being made to the participant. The joint and survivor annuity must be actuarially equivalent to a single life annuity for the life of the participant.

The plan is not required to absorb the cost for either benefit. This cost may be passed along to participants and their spouses, typically by reducing the retirement benefit. At any time the participant is allowed to waive the QJSA form of benefit, the QPSA form of benefit, or both, and he or she is allowed to revoke his or her selection at any time during the election period. The spouse must consent to that election. The spouse's consent must be in writing and must be witnessed by a plan representative or a notary public. These annuities need not be provided if the participant and his or her spouse have been married less than one year (Allen et al., 2003 and Boyers, 1986).

### 1.3 Personal Savings

The third pillar in the United States is personal savings (including individual insurance and annuities). *Annuities* are periodic payments for a fixed period of time or for the duration of a designated life or lives. The person who receives the periodic payment is called the *annuitant*. There are different types of annuities. We will focus on *Joint Life Annuities* and *Joint-and Last-Survivor-Annuities*. Joint life annuities for two lives provide payments as long as both persons are alive. Benefit payments cease upon the first death. Such a plan is appropriate only when two people have another source of income that is sufficient for one person, but not for both. These contracts are not popular. Joint-and-last-survivor annuities pay benefits based on the lives of two or more persons, such as a husband and wife. The insurer pays as long as either of the annuitants is alive. Usually the benefits are paid for longer periods of time than under single life annuities. That is why joint-and-last-survivor annuities are more expensive than single life annuities. In spite the higher cost, this kind of annuity is attractive for many couples who need an income as long as either is alive. There are two features to keep the cost of this annuity reasonably low. First, it does not need any guarantee period because benefits will continue for the surviving spouse. Second, many annuities pay only two-thirds or one-half of the original income after the first death.

For married couples, when both persons die, the children and perhaps grandchildren have to pay the federal estate tax. In the case of a larger estate, this may be a huge amount of money and the children or grandchildren may be forced to sell in order to pay the estate tax all or part of the estate. To conserve the size of a larger estate after their death, a couple can buy *Joint Survivorship Life Insurance*. The policy covers two lives as the insureds in a single policy. The death benefits are payable to the beneficiary at the death of the second insured. Such an insurance is not appropriate to meet family income needs after the death of the first insured. The premium for joint survivorship life insurance is usually significantly less than the premium for comparable individual life insurance. This occurs because the policy covers two lives and does not pay until the second death (Hallmann, 1994).

In all cases mentioned above - QPSA, QJSA, joint-life-annuities, joint-and last-survivor-annuities and joint survivorship life insurance - the payments are based on a combination of two lives. My thesis discusses modelling such combinations of lives.

## CHAPTER II

## ACTUARIAL NOTATION

### 2.1 Single Life Functions

We will adopt the notation used in Bowers et al. (1997). Let $X$ be the age-at-death random variable of a newborn and let $F_X(x)$ denote its distribution function,

$$F_X(x) = Pr(X \leq x),\ x \geq 0.$$

Let $f_X(x)$ denote its density function. We have the relationship

$$F'_X(x) = f_X(x).$$

The *survival function* is denoted by

$$s(x) = 1 - F_X(x)$$

The symbol $(x)$ denotes a person aged $x$. Let $T(x)$ be the future lifetime of $(x)$, $T(x) = X - x$. Let ${}_tq_x = Pr(T(x) \leq t)$ and ${}_tp_x = Pr(T(x) \geq t)$ be the probability that $(x)$ will die within $t$ years and the probability that $(x)$ will survive the next $t$ years, respectively.

**Definition 1** *The greatest integer function is determined by the equation $y = int(x)$, where the value of $y$ that corresponds to $x$ is the greatest integer that is less than or* *equal to $x$.*

Let $K(x)$ be the *curtate-future-lifetime*; meaning the number of future years completed by $(x)$ prior to death. $K(x)$ is the greatest integer function of $T(x)$. The probability that a person aged $x$ lives exactly $k$ years is:

$$\begin{aligned}
Pr(K = k) &= Pr(k \leq T(x) < k+1) \\
&= Pr(T(x) < k+1) - Pr(T(x) < k) \\
&= (1 - {}_{k+1}p_x) - (1 - {}_k p_x) \\
&= {}_k p_x - {}_{k+1}p_x \\
&= {}_k p_x - {}_k p_x \cdot p_{x+k} \\
&= {}_k p_x (1 - p_{x+k}) \\
&= {}_k p_x q_{x+k} \qquad (1)
\end{aligned}$$

**Definition 2** *The force of mortality $\mu(x)$ is defined by*

$$\mu(x) = \frac{f_X(x)}{1 - F_X(x)}$$

*It gives the value of the conditional p.d.f. of $X$ at exact age $x$, given survival to that age.*

Since $s'(x) = -f_X(x)$, we have

$$\mu(x) = \frac{f_X(x)}{1 - F_X(x)} = \frac{-s'(x)}{s(x)} = -\frac{d}{dx}\log[s(x)]$$

Integrating this expression from $x$ to $x+n$, we have

$$\begin{aligned}
& -\int_x^{x+n} \mu(y)dy = \int_x^{x+n} \frac{d}{dy}\log[s(y)]dy \\
\Leftrightarrow\ & -\int_x^{x+n} \mu(y)dy = \log[s(x+n)] - \log[s(x)] \\
=\ & \log\left[\frac{s(x+n)}{s(x)}\right] = \log[{}_np_x] \\
\Leftrightarrow\ & {}_np_x = \exp\left[-\int_x^{x+n} \mu(y)dy\right].
\end{aligned}$$

Substituting $x+s$ by $y$, we have

$${}_np_x = \exp\left[-\int_0^n \mu(x+s)ds\right]. \tag{2}$$

Let $F_{T(x)}(t)$ and $f_{T(x)}(t)$ denote the distribution and density function of $T(x)$, respectively. Since $F_{T(x)}(t) = {}_tq_x$, we have

$$\begin{aligned}
f_{T(x)}(t) &= \frac{d}{dt}{}_tq_x = \frac{d}{dt}\left[1 - \frac{s(x+t)}{s(x)}\right] = -\frac{s'(x+t)}{s(x)} \\
&= \frac{s(x+t)}{s(x)} \cdot \frac{-s'(x+t)}{s(x+t)} \quad = {}_tp_x \cdot \mu(x+t),\ t \geq 0.
\end{aligned} \tag{3}$$

**Definition 3** *The complete-expectation-of-life is defined as*

$$E[T(x)] = \int_0^\infty t \cdot {}_tp_x \cdot \mu(x+t)dt$$

*and denoted by* $\mathring{e}_x$*, assuming that the expected value exists.*

Using integration by parts we get

$$\mathring{e}_x = \int_0^\infty t \cdot \frac{d}{dt}F_{T(x)}(t)dt = \int_0^\infty t \cdot \frac{d}{dt}(-{}_tp_x)dt$$

$$= \quad t(-{}_tp_x)\Big|_{t=0}^{\infty} + \int_0^{\infty} {}_tp_x dt$$

**Lemma 1** *When we assume that the expected value of $T(x)$ exists, we have*

$$\lim_{t\to\infty} t(-{}_tp_x) = 0.$$

The proof is based on an idea by Fisz (1963):

**Proof:**

$$\begin{aligned}
& \lim_{t\to\infty} t({}_tp_x) \\
= \quad & \lim_{t\to\infty} t \cdot Pr(T(x) > t) \\
= \quad & \lim_{t\to\infty} t \cdot \int_t^{\infty} f_{T(x)}(u)du \\
= \quad & \lim_{t\to\infty} \int_t^{\infty} t \cdot f_{T(x)}(u)du \\
\leq \quad & \lim_{t\to\infty} \int_t^{\infty} u \cdot f_{T(x)}(u)du \\
= \quad & 0.
\end{aligned}$$

Since

$$\lim_{t\to\infty} t({}_tp_x) \geq 0$$

and

$$\lim_{t\to\infty} t({}_tp_x) \leq 0$$

we have

$$\lim_{t\to\infty} t({}_tp_x) = 0. \tag{4}$$

Thus, we have

$$\lim_{t\to\infty} t(-{}_tp_x) = -\lim_{t\to\infty} t({}_tp_x) = 0.$$

$\square$

When we assume that the expected value of $T(x)$ exists, we have by Lemma 1

$$\mathring{e}_x = \int_0^\infty {}_tp_x dt \tag{5}$$

The second moment of the future lifetime is

$$\begin{aligned} E[T^2] &= \int_0^\infty t^2 \cdot {}_tp_x \cdot \mu(x+t)dt = \int_0^\infty t^2 \frac{d}{dt}(-{}_tp_x)dt \\ &= t^2(-{}_tp_x)\Big|_{t=0}^\infty + \int_0^\infty 2t{}_tp_x dt \end{aligned}$$

We assume that $E[T^2(x)]$ exists. Thus, we have

$$E[T^2] = 2\int_0^\infty t{}_tp_x dt \tag{6}$$

And the variance of the future lifetime is

$$\begin{aligned} Var[T] &= E[T^2] - E^2[T] \\ &= 2\int_0^\infty t{}_tp_x dt - \mathring{e}_x^2 \end{aligned} \tag{7}$$

**Definition 4** *Assuming that the expected value of $K(x)$ exists, the curtate-expectation-of-life is defined as*

$$E[K(x)] = \sum_{k=0}^{\infty} k\,{}_kp_x q_{x+k} = \sum_{k=0}^{\infty} {}_kp_x$$

*and denoted by $e_x$.*

Following Bowers et al. (1997) we will now discuss various forms of annuities. A *life annuity* is a series of payments made continuously or at equal intervals (such as months, quarters, years) until a given life dies. It may be temporary, meaning limited to a given number of years, or it may be payable for the whole life. The payments may commence immediately, or the annuity may be deferred. Payments may be due at the beginnings of the payment intervals (*annuities-due*) or at the end of such intervals (*annuities-immediate*). We assume a constant effective annual rate of interest $i$ (or the equivalent constant force of interest $\delta$). The *discount factor* is denoted as $v$ and is equal to $\frac{1}{1+i}$. The *present value* of the amount $C$ paid at time $t$ depends on the discount factor $v$ and is equal to $v^tC$. We start with annuities payable continuously at the rate of 1 per year. A *whole life annuity* provides for payments until death. Let $\bar{a}_{\overline{n|}}$ denote the present value of a level annual payment of 1 paid continuously.

$$\bar{a}_{\overline{n|}} = \int_0^n v^t dt = \int_0^n e^{-\delta t} dt = \frac{1}{\delta} - \frac{1}{\delta} e^{-\delta n},$$

Hence, the present value of payments to be made is $Y = \bar{a}_{\overline{T|}}$ for all $T \geq 0$ where $T$ is the future lifetime of $(x)$. The expected present value, called the actuarial present value, for a continuous whole life annuity is denoted by $\bar{a}_x$ where the past fixed subscript, $x$, indicates that the annuity ceases when $(x)$ dies. As shown in (3), the p.d.f. of $T$ is ${}_tp_x\mu(x+t)$ and the actuarial present value can be calculated by

$$\bar{a}_x = E[Y] = E[\bar{a}_{\overline{T|}}] = \int_0^\infty \bar{a}_{\overline{t|}} \cdot {}_tp_x\mu(x+t)dt \tag{8}$$
$$= \int_0^\infty \bar{a}_{\overline{t|}} \frac{d}{dt}(-{}_tp_x)dt$$

We integrate by parts with $f(t) = \bar{a}_{\overline{t|}}$ and $g'(t) = {}_tp_x\mu(x+t) = \frac{d}{dt}[F_T(t)] = \frac{d}{dt}[1-{}_tp_x]$, implying that $f'(t) = (\frac{1}{\delta} - \frac{1}{\delta}\exp(-\delta t))' = \exp(-\delta t) = v^t$ and that $g(t) = -{}_tp_x$. So, we get:

$$\bar{a}_x = \bar{a}_{\overline{t|}} \cdot (-{}_tp_x)\Big|_{t=0}^{\infty} + \int_0^\infty v^t{}_tp_x dt$$

Since $E[T^2] < \infty$, we have

$$\bar{a}_x = \int_0^\infty v^t{}_tp_x dt \tag{9}$$

We now turn to temporary and deferred life annuities. An *n-year temporary life annuity* pays continuously while $(x)$ survives during the next $n$ years. The present value of a benefits random variable for such an annuity of 1 per year is

$$Y = \begin{cases} \bar{a}_{\overline{T|}} & 0 \le T < n \\ \bar{a}_{\overline{n|}} & T \ge n \end{cases}$$

The expected present value, called the actuarial present value, of an n-year temporary life annuity is denoted by $\bar{a}_{x:\overline{n|}}$ and equals

$$\bar{a}_{x:\overline{n|}} = E[Y] = \int_0^n \bar{a}_{\overline{t|}} \cdot {}_tp_x\mu(x+t)dt + \bar{a}_{\overline{n|}} \cdot {}_np_x$$

Integrating by parts with $f(t) = \bar{a}_{\overline{t|}}$ and $g'(t) = {}_tp_x\mu(x+t)$, implying that $f'(t) = v^t$ and $g(t) = -{}_tp_x$, we have

$$\bar{a}_{x:\overline{n|}} = \bar{a}_{\overline{t|}} \cdot (-{}_tp_x)\Big|_{t=0}^{n} + \int_0^n v^t {}_tp_x dt + \bar{a}_{\overline{n|}} \cdot {}_np_x = \int_0^n v^t {}_tp_x dt \tag{10}$$

The analysis for an *n-year deferred whole life annuity* is similar. This annuity commences its payments $n$ years after the policy becomes effective and then as long as $(x)$ survives. The present value random variable $Y$ is defined as

$$Y = \begin{cases} 0 = \bar{a}_{\overline{T|}} - \bar{a}_{\overline{T|}} & 0 \le T < n \\ v^n \bar{a}_{\overline{T-n|}} = \bar{a}_{\overline{T|}} - \bar{a}_{\overline{n|}} & T \ge n \end{cases}$$

**Lemma 2**

$$v^n \bar{a}_{\overline{T-n|}} = \bar{a}_{\overline{T|}} - \bar{a}_{\overline{n|}}$$

**Proof**:

$$v^n \bar{a}_{\overline{T-n|}} = v^n \int_0^{T-n} v^t dt = \int_0^{T-n} v^{t+n} dt$$

Substituting $u$ by $t+n$, gives

$$v^n \bar{a}_{\overline{T-n|}} = \int_n^T v^u du = \int_0^T v^u du - \int_0^n v^u du = \bar{a}_{\overline{T|}} - \bar{a}_{\overline{n|}}. \qquad \square$$

Note that, from the definitions of $Y$,

$$\begin{aligned} & (Y \text{ for an n-year deferred whole life annuity}) \\ = \; & (Y \text{ for a whole life annuity}) - (Y \text{ for an } n\text{-year temporary life annuity}). \end{aligned}$$

Hence, the actuarial present value of an $n$-year deferred whole life annuity, denoted by ${}_{n|}\bar{a}_x$, is

$$ {}_{n|}\bar{a}_x = \bar{a}_x - \bar{a}_{x:\overline{n|}} $$

We now turn to the analysis of an $n$-year certain and life annuity. This is a whole life annuity with a guarantee of payments for the first $n$ years. The present value of annuity payments is

$$ Y = \begin{cases} \bar{a}_{\overline{n|}} & T \le n \\ \bar{a}_{\overline{T|}} & T > n \end{cases} $$

The actuarial present value is denoted by $\bar{a}_{\overline{x:\overline{n|}}}$.

$$ \bar{a}_{\overline{x:\overline{n|}}} = E(Y) = \int_0^n \bar{a}_{\overline{n|}} \cdot {}_tp_x\mu(x+t)dt + \int_n^\infty \bar{a}_{\overline{t|}} \cdot {}_tp_x\mu(x+t)dt $$
$$ = \; {}_nq_x \cdot \bar{a}_{\overline{n|}} + \int_n^\infty \bar{a}_{\overline{t|}} \frac{d}{dt}(-{}_tp_x)dt. $$

Using integration-by-parts with $f(t) = \bar{a}_{\overline{t|}}$, $g'(t) = {}_tp_x\mu(x+t)$, $f'(t) = v^t$, $g(t) = -{}_tp_x$, we have

$$ \bar{a}_{\overline{x:\overline{n|}}} = {}_nq_x \cdot \bar{a}_{\overline{n|}} + \bar{a}_{\overline{t|}}(-{}_tp_x)\Big|_n^\infty + \int_n^\infty v^t{}_tp_xdt. $$

Because the expected value of the future lifetime is finite, we have

$$ {}_nq_x \cdot \bar{a}_{\overline{n|}} + \bar{a}_{\overline{n|}} \cdot {}_np_x + \int_n^\infty v^t{}_tp_xdt = \bar{a}_{\overline{n|}}({}_nq_x + {}_np_x) + \int_n^\infty v^t{}_tp_xdt $$
$$ = \; \bar{a}_{\overline{n|}} + \int_n^\infty v^t{}_tp_xdt $$

The theory of discrete life annuities is analogous to the theory of continuous life annuities, with integrals replaced by sums and integrands replaced by summands. For continuous annuities there was no distinction between payments at the beginning of payment intervals or at the ends, meaning, between annuities-due and annuities-immediate. For discrete annuities, we need this distinction, and we start with annuities-due. A *whole-life annuity-due* is an annuity that pays a unit amount at the beginning of each year that the annuitant $(x)$ survives. Let $\ddot{a}_{\overline{n|}}$ denote the present value of a level annual payment of 1 dollar paid at the beginning of each year of $n$ years.

$$\ddot{a}_{\overline{n|}} = \sum_{k=0}^{n-1} v^k = \frac{1-v^n}{1-v}.$$

The present value random variable, $Y$, for a whole life annuity-due is $Y = \ddot{a}_{\overline{K+1|}}$, where $K$ is the curtate-future lifetime of $(x)$. The actuarial present value of a whole life annuity-due can be calculated by

$$\ddot{a}_x = E(Y) = \sum_{k=0}^{\infty} \ddot{a}_{\overline{k+1|}} \cdot {}_kp_x \cdot q_{x+k} \tag{11}$$

To simplify this we need summation by parts (Seton Hall University, 2003):

**Lemma 3** *Consider the sequences* $\{a_n\}_{n=1}^{\infty}$ *and* $\{b_n\}_{n=1}^{\infty}$*. Let* $S_N = \sum_{n=1}^{N} a_n$ *be the* $n$*-th partial sum. Then for any* $0 < m \leq n$ *we have*

$$\sum_{j=m}^{n-1} S_j(b_j - b_{j+1}) = \sum_{j=m}^{n} a_j \cdot b_j - [S_n \cdot b_n - S_{m-1} \cdot b_m]$$

**Proof:**

$$\begin{aligned}
\sum_{j=m}^{n} a_j \cdot b_j &= \sum_{j=m}^{n} (S_j - S_{j-1}) \cdot b_j \\
&= \sum_{j=m}^{n} S_j \cdot b_j - \sum_{j=m}^{n} S_{j-1} \cdot b_j \\
&= \sum_{j=m}^{n} S_j \cdot b_j - \sum_{j=m-1}^{n-1} S_j \cdot b_{j+1} \\
&= \sum_{j=m}^{n-1} S_j \cdot (b_j - b_{j+1}) + S_n b_n - S_{m-1} b_m .
\end{aligned}$$

□

If $n = \infty$ the formula reduces to

$$\sum_{j=m}^{\infty} S_j (b_j - b_{j+1}) = \sum_{j=m}^{\infty} a_j \cdot b_j + S_{m-1} \cdot b_m \tag{12}$$

**Proof:**

$$\begin{aligned}
\sum_{j=m}^{\infty} a_j \cdot b_j &= \sum_{j=m}^{\infty} (S_j - S_{j-1}) \cdot b_j \\
&= \sum_{j=m}^{\infty} S_j \cdot b_j - \sum_{j=m}^{\infty} S_{j-1} \cdot b_j \\
&= \sum_{j=m}^{\infty} S_j \cdot b_j - \sum_{j=m-1}^{\infty} S_j \cdot b_{j+1} \\
&= \sum_{j=m}^{\infty} S_j \cdot (b_j - b_{j+1}) - S_{m-1} b_m .
\end{aligned}$$

□

Note that $\ddot{a}_{\overline{K+1|}} = \sum_{k=0}^{K} v^k$ and ${}_k p_x \cdot q_{x+k} = {}_k p_x \cdot (1 - p_{x+k}) = {}_k p_x - {}_{k+1} p_x$. We choose

$S_j = \ddot{a}_{\overline{j+1|}}$, $b_j = {}_jp_x$, $a_j = v^j$ and $m = 1$. Now we can use formula (12) to obtain

$$\begin{aligned}
\ddot{a}_x &= \sum_{k=0}^{\infty} \ddot{a}_{\overline{k+1|}} \cdot {}_kp_x \cdot q_{x+k} = \sum_{k=0}^{\infty} S_k(b_k - b_{k+1}) \\
&= S_0(b_0 - b_1) + \sum_{k=1}^{\infty} S_k(b_k - b_{k+1}) \\
&= \ddot{a}_{\overline{1|}}({}_0p_x - p_x) + \sum_{k=1}^{\infty} v^k {}_kp_x + S_0 b_1 = q_x + \sum_{k=1}^{\infty} v^k {}_kp_x + \sum_{j=0}^{0} v^j \cdot p_x \\
&= q_x + \sum_{k=1}^{\infty} v^k {}_kp_x + p_x = 1 + \sum_{k=1}^{\infty} v^k {}_kp_x = \sum_{k=0}^{\infty} v^k {}_kp_x \qquad (13)
\end{aligned}$$

The present-value random variable of an $n$-year temporary life annuity-due of 1 per year is

$$Y = \begin{cases} \ddot{a}_{\overline{K+1|}} & 0 \le K < n \\ \ddot{a}_{\overline{n|}} & K \ge n \end{cases}$$

and its actuarial present value is

$$\begin{aligned}
\ddot{a}_{x:\overline{n|}} &= E(Y) = \sum_{k=0}^{n-1} \ddot{a}_{\overline{k+1|}} \cdot {}_kp_x \cdot q_{x+k} + \ddot{a}_{\overline{n|}} \cdot {}_np_x \\
&= \ddot{a}_{\overline{1|}} \cdot {}_0p_x \cdot q_x + \sum_{k=1}^{n-1} \ddot{a}_{\overline{k+1|}} \cdot {}_kp_x \cdot q_{x+k} + \ddot{a}_{\overline{n|}} \cdot {}_np_x \\
&= q_x + \sum_{k=1}^{n-1} \ddot{a}_{\overline{k+1|}} \cdot {}_kp_x \cdot q_{x+k} + \ddot{a}_{\overline{n|}} \cdot {}_np_x
\end{aligned}$$

We use Lemma 3 with $S_j = \ddot{a}_{\overline{j+1|}}$, $b_j = {}_jp_x$, $a_j = v^j$, $m = 1$ and $n = n$ to obtain

$$\ddot{a}_{x:\overline{n|}} = q_x + \ddot{a}_{\overline{n|}} \cdot {}_np_x + \sum_{k=1}^{n} v^k {}_kp_x - [\ddot{a}_{\overline{n+1|}} \cdot {}_np_x - \ddot{a}_{\overline{1|}} \cdot p_x]$$

$$
\begin{aligned}
&= q_x + \ddot{a}_{\overline{n|}} \cdot {}_np_x + \sum_{k=1}^{n-1} v^k {}_kp_x + v^n {}_np_x - \ddot{a}_{\overline{n+1|}} {}_np_x + p_x \\
&= 1 + \ddot{a}_{\overline{n|}} \cdot {}_np_x + \sum_{k=1}^{n-1} v^k {}_kp_x + {}_np_x(v^n - \ddot{a}_{\overline{n+1|}}) \\
&= 1 + \ddot{a}_{\overline{n|}} \cdot {}_np_x + \sum_{k=1}^{n-1} v^k {}_kp_x + {}_np_x(v^n - \sum_{j=0}^{n} v^j) \\
&= 1 + \ddot{a}_{\overline{n|}} \cdot {}_np_x + \sum_{k=1}^{n-1} v^k {}_kp_x + {}_np_x(- \sum_{j=0}^{n-1} v^j) \\
&= 1 + \ddot{a}_{\overline{n|}} \cdot {}_np_x + \sum_{k=1}^{n-1} v^k {}_kp_x - {}_np_x \cdot \ddot{a}_{\overline{n|}} = 1 + \sum_{k=1}^{n-1} v^k {}_kp_x = \sum_{k=0}^{n-1} v^k {}_kp_x \qquad (14)
\end{aligned}
$$

For an *n-year deferred whole life annuity-due* of 1 payable at the beginning of each year while $(x)$ survives from $x+n$ onward, the present-value random variable is

$$
Y = \begin{cases} 0 & 0 \le K < n \\ v^n \ddot{a}_{\overline{K+1-n|}} & K \ge n \end{cases}
$$

and its actuarial present value is

$$
E(Y) = {}_{n|}\ddot{a}_x = \sum_{k=n}^{\infty} v^n \ddot{a}_{\overline{K+1-n|}} \cdot {}_kp_x \cdot q_{x+k} = \sum_{k=n}^{\infty} v^k {}_kp_x
$$

The last equality follows again by summation-by-parts. The procedures above for annuities-due can be adapted for annuities immediate. Payments for this kind of annuity are made at the ends of the payment periods. For example, for a *whole life annuity-immediate*, the present value random variable is

$$
Y = a_{\overline{K|}},
$$

where

$$a_{\overline{K|}} = \sum_{j=1}^{K} v^j.$$

Then,

$$a_x = E(Y) = \sum_{k=0}^{\infty} {}_kp_x \cdot q_{x+k} \cdot a_{\overline{k|}} = \sum_{k=0}^{\infty} {}_kp_x \cdot q_{x+k}(\ddot{a}_{\overline{k+1|}} - 1)$$

$$= \ddot{a}_x - \sum_{k=0}^{\infty} {}_kp_x \cdot q_{x+k} = \ddot{a}_x - 1 = \sum_{k=1}^{\infty} {}_kp_x \cdot v^k$$

2.2 Multiple Life Functions

In this subsection we want to discuss density functions, probability functions, force of mortality and annuities for two lives. We follow Bowers et at. (1997). A useful definition is that of *status* for which there are definitions of survival and failure. In order to define a status we need two elements. Since there is a broad range of application of the concept, the general term *entities* is used in the definition.

· There must be a finite set of entities; and for each member it must be possible to define a future lifetime random variable.

· It must be possible to determine the survival of the status at any future time.

To illustrate the meaning of a status let's look at an example: A single life $(x)$ defines a status that survives while $(x)$ is alive. Thus the random variable $T(x)$, used in the previous subsection to denote the future lifetime of $(x)$, can be interpreted as the period of survival of the status and also as the time-until-failure of the status. The time-until-failure of a status is a function of the future lifetimes of the lives involved. In theory these future lifetimes will be dependent. The joint distribution function of

$T(x)$ and $T(y)$ is

$$F_{T(x)T(y)}(s,t) = Pr(T(x) \leq s, T(y) \leq t) = \int_{-\infty}^{s} \int_{-\infty}^{t} f_{T(x)T(y)}(u,v)dvdu$$

and the joint survival function of $T(x)$ and $T(y)$ is

$$s_{T(x)T(y)}(s,t) = Pr(T(x) > s, T(y) > t) = \int_{s}^{\infty} \int_{t}^{\infty} f_{T(x)T(y)}(u,v)dvdu$$

2.2.1 The Joint-Life-Status

The *joint-life-status* survives while every member of a set of lives is alive and fails when the first member dies. It is denoted by $(x_1, ..., x_m)$, where $x_i$ is the age of member $i$ and $m$ is the number of members. Notation introduced in the previous subsection is used here with the subscript listing several ages rather than a single age. For example, $\ddot{a}_{xy}$ and ${}_tp_{xy}$ have the same meaning for the joint-life status $(xy)$ as $\ddot{a}_x$ and ${}_tp_x$ have for the single life $(x)$.

We consider the distribution of the time-until-failure of a joint-life status. For $m$ lives, $T(x_1, ...., x_m) = \min[T(x_1), ..., T(x_m)]$, where $T(x_i)$ is the future lifetime of individual $i$. In the special case of two lives, $(x)$and $(y)$, we have $T(xy) = \min[T(x), T(y)]$. When indicated by context, we denote the future lifetime of the joint-life status by simply $T$. The distribution function of $T$, for $t > 0$, in terms of the joint distribution of $T(x)$ and $T(y)$ is

$$\begin{aligned} & F_T(t) = {}_tq_{xy} = Pr(T \leq t) = Pr[\min(T(x), T(y)) \leq t] \\ = \;& 1 - Pr[\min(T(x), T(y)) > t] = 1 - Pr[T(x) > t \,\text{and}\, T(y) > t] \\ = \;& 1 - s_{T(x)T(y)}(t,t). \end{aligned} \tag{15}$$

Note that

$$_tp_{xy} = Pr[T(xy) > t] = 1 - F_{T(xy)}(t) = s_{T(x)T(y)}(t,t)$$

As explained in the previous subsection, the distribution of $T$ can be specified by the force of mortality. The traditional notation for this force is $\mu_{x+t:\,y+t}$ (analogous to $\mu_{x+t}$) but we use the notation $\mu_{xy}(t)$. By analogy with the first formula (2) and with $f_{T(x)}(x)$ and $F_{T(x)}(x)$ replaced by $f_{T(xy)}(t)$ and $F_{T(xy)}(t)$, we have

$$\mu_{xy}(t) = \frac{f_{T(xy)}(t)}{1 - F_{T(xy)}(t)} \tag{16}$$

**Theorem 1** *If $T(x)$ and $T(y)$ are independent the following conditions hold:*

*(1)* $_tp_{xy} = {}_tp_x \cdot {}_tp_y$

*(2)* $\mu_{xy}(t) = \mu(x+t) + \mu(y+t)$

**Remark 1** *(2) means that if the future lifetimes are independent, the force of mortality for their joint-life status is the sum of the forces of mortality of the individuals.*

**Proof:**

First, we prove property (1) of Theorem 1: Since $T(x)$ and $T(y)$ are independent we have

$$_tp_{xy} = Pr(T(x) > t, T(y) > t) = Pr(T(x) > t) \cdot Pr(T(y) > t) = {}_tp_x \cdot {}_tp_y$$

Second, we want to show property (2) of Theorem 1: Note that

$$s_{T(x)T(y)}(t,t) = \int_t^\infty \int_t^\infty f_{T(x)T(y)}(u,v)dudv.$$

We want to find the derivative of $s_{T(x)T(y)}(t,t)$ with respect to $t$. To find this derivative, we need the *Leibniz's rule*, as in Zwillinger (1992):

$$\begin{aligned} & \frac{d[\int_{\alpha(t)}^{\beta(t)} g(v,t)dv]}{dt} \\ = & \int_{\alpha(t)}^{\beta(t)} \frac{d}{dt} g(v,t)dv + g(\beta(t),t)\frac{d}{dt}\beta(t) - g(\alpha(t),t)\frac{d}{dt}\alpha(t) \end{aligned} \tag{17}$$

In our case

$$\alpha(t) = t, \beta(t) = \infty, g(v,t) = \int_t^\infty f_{T(x)T(y)}(u,v)du,$$

Thus, using (17), we have

$$\begin{aligned} \frac{d}{dt} g(v,t) &= \frac{d}{dt}\int_t^\infty f_{T(x)T(y)}(u,v)du \\ &= \int_t^\infty 0 + 0 - f_{T(x)T(y)}(t,v)\cdot 1 \\ &= -f_{T(x)T(y)}(t,v). \end{aligned}$$

Using (17), this implies that

$$\begin{aligned} & \frac{d}{dt} s_{T(x)T(y)}(t,t) = \frac{d}{dt}\int_t^\infty g(v,t)dv \\ = & \int_t^\infty -f_{T(x)T(y)}(t,v)dv + 0 - g(t,t)\cdot 1 \\ = & -\int_t^\infty f_{T(x)T(y)}(t,v)dv - \int_t^\infty f_{T(x)T(y)}(u,t)du. \end{aligned} \tag{18}$$

Using (15) and (18) we have

$$f_{T(xy)}(t) = \frac{d}{dt}F_{T(xy)}(t) = \frac{d}{dt}[1 - s_{T(x)T(y)(t,t)}] = -\frac{d}{dt}s_{T(x)T(y)}(t,t)$$

$$= \int_t^\infty f_{T(x)T(y)}(t,v)dv + \int_t^\infty f_{T(x)T(y)}(u,t)du. \tag{19}$$

Since $T(x)$ and $T(y)$ are independent and using (3) and (19), we have

$$f_{T(x)T(y)}(u,v) = f_{T(x)}(u) \cdot f_{T(y)}(v) = {}_u p_x \cdot \mu(x+u) \cdot {}_v p_y \cdot \mu(y+v)$$

and

$$\begin{aligned} f_{T(xy)}(t) &= \int_t^\infty {}_t p_x \cdot \mu(x+t) \cdot {}_v p_y \cdot \mu(y+v)dv \\ + & \int_t^\infty {}_u p_x \cdot \mu(x+u) \cdot {}_t p_y \cdot \mu(y+t)du \\ = & {}_t p_x \cdot \mu(x+t) \int_t^\infty f_{T(y)}(v)dv + {}_t p_y \cdot \mu(y+t) \int_t^\infty f_{T(x)}(u)du \\ = & {}_t p_x \cdot \mu(x+t) {}_t p_y + {}_t p_y \cdot \mu(y+t) {}_t p_x = {}_t p_x \cdot {}_t p_y [\mu(x+t) + \mu(y+t)]. \end{aligned} \tag{20}$$

Thus, using (20) and (16), we have

$$\begin{aligned} \mu_{xy}(t) = \frac{f_{T(xy)}(t)}{1 - F_{T(xy)}(t)} &= \frac{{}_t p_x \cdot {}_t p_y [\mu(x+t) + \mu(y+t)]}{1 - [1 - s_{T(x)T(y)}(t,t)]} \\ = \frac{{}_t p_x \cdot {}_t p_y [\mu(x+t) + \mu(y+t)]}{{}_t p_{xy}} &= \frac{{}_t p_x \cdot {}_t p_y [\mu(x+t) + \mu(y+t)]}{{}_t p_x \cdot {}_t p_y} \\ = \mu(x+t) + \mu(y+t). \end{aligned}$$

□

Let's consider an annuity payable continuously at the rate of 1 dollar per year until $(u)$ fails, the present value random variable for such an annuity is $Y = \bar{a}_{\overline{T|}}$, where $T$ is the future lifetime of $(u)$. The actuarial present value of the annuity, $\bar{a}_u$, is: (compare

(8) and (9))

$$\bar{a}_u = \int_0^\infty \bar{a}_{\overline{t|}} \cdot {}_tp_u\mu(u+t)dt = \int_0^\infty {}_tp_u \cdot v^t dt$$

We want to apply this relationship to an annuity payable continuously at the rate of 1 dollar per year while both persons $(x)$ and $(y)$ are alive. This is an annuity in respect to $(xy)$. Thus, we have:

$$\bar{a}_{xy} = \int_0^\infty {}_tp_{xy} \cdot v^t dt$$

2.2.2 The Last-Survivor Status

The last-survivor status exists as long as at least one member of a set of lives is alive and fails when the last member dies. It is denoted by $(\overline{x_1, ..., x_m})$, where - as before - $x_i$ is the age of member $i$ and $m$ is the number of the members. The future lifetime of the last-survivor is denoted by $T(\overline{x_1, ..., x_m})$, and it is equal to $T(\overline{x_1, ..., x_m}) = \max[T(x_1), ..., T(x_m)]$, where $T(x_i)$ is the future lifetime of member $i$. We only consider the case of two lives $(x)$ and $(y)$. The future lifetime of the joint-life status in this case is $T(\overline{xy}) = \max[T(x), T(y)]$. The distribution function of $T(\overline{xy})$ is

$$\begin{aligned} & F_{T(\overline{xy})}(t) = Pr[T(\overline{xy}) \leq t] = Pr[\max(T(x), T(y)) \leq t] \\ = \; & Pr[T(x) \leq t \,\text{and}\, T(y) \leq t] = 1 - {}_tp_{\overline{xy}} \end{aligned}$$

and

$$f_{T(\overline{xy})}(t) = \frac{d}{dt} F_{T(\overline{xy})}(t)$$

There are relationships among $T(xy)$, $T(\overline{xy})$, $T(x)$ and $T(y)$:

$T(xy)$ is either equal to $T(x)$ or equal to $T(y)$. $T(\overline{xy})$ is equal to the other one. That's why the following equations hold:

$$T(xy) + T(\overline{xy}) = T(x) + T(y)$$

$$T(xy)T(\overline{xy}) = T(x)T(y)$$

From probability, we know that

$$Pr(A \cup B) + Pr(A \cap B) = Pr(A) + Pr(B).$$

If $A = \{T(x) \leq t\}$ and $B = \{T(y) \leq t\}$, we have $A \cup B = \{T(xy) \leq t\}$ and $A \cap B = \{T(\overline{xy}) \leq t\}$ and then

$$F_{T(xy)}(t) + F_{T(\overline{xy})}(t) = F_{T(x)}(t) + F_{T(y)}(t).$$

This implies

$${}_tp_{xy} + {}_tp_{\overline{xy}} = {}_tp_x + {}_tp_y$$

# CHAPTER III

# GOMPERTZ LAW

## 3.1 Single Lives

There are two main justifications for postulating an analytic law for mortality. Those justifications are from Higgins (2003) and Bowers et al. (1997):

*1. Philosophic:* Since many phenomena which are observed in physics are governed by simple formulas, some authors have suggested that human mortality can be explained by a simple law with biological arguments.

*2. Practical:* It is more convenient to operate with a function with only a few parameters than with a life table with perhaps 100 parameters. Besides, it is easier to estimate functions like life expectations, conditional probabilities of survival, etc. Since some analytic forms have elegant properties it is convenient to evaluate probabilities for more than one life.

The earliest model still in use and the most influential parametric mortality model in literature is that of Benjamin Gompertz from 1825. He recognized that the behavior of human mortality for large portions of the life table is exponential. The original Gompertz law is

$$\mu(x) = Bc^x, \tag{21}$$

where $B$ and $c$ are positive unknown parameters or

$$\mu(x) = Be^{bx}, \tag{22}$$

where $\mu(x)$ is the force of mortality at age $x$ as defined in chapter 2. If we compare those two equations, we can see that one is only the reparametrization of the other: Substituting $e^b$ for $c$ in (21), we get (22). This implies that

$$\mu(x+t) = Be^{b(x+t)} = Be^{bx}e^{bt} = \mu(x)e^{bt}. \tag{23}$$

Following Gajek and Ostaszewski (2002, 2003) we want to calculate the expected value, the variance and the variability coefficient of the future lifetime $T$ under Gompertz' law of mortality. Using (2), the probability of surviving $t$ years for a life $(x)$ under Gompertz' law of mortality is

$$\begin{aligned} {}_tp_x &= \exp\left[-\int_0^t \mu(x+s)ds\right] = \exp\left[-\int_0^t \mu(x)e^{bs}ds\right] \\ &= \exp\left[-\frac{\mu(x)}{b}e^{bs}\Big|_{s=0}^{t}\right] = \exp\left[-\frac{\mu(x)}{b}e^{bt} + \frac{\mu(x)}{b}\right] = \exp\left[-\frac{\mu(x)}{b}(e^{bt}-1)\right] \end{aligned}$$

Since $f_T(t) = {}_tp_x\mu(x+t)$ by (3), we have that

$$f_T(t) = \exp\left[-\frac{\mu(x)}{b}(e^{bt}-1)\right] \cdot \mu(x) \cdot e^{bt} = \mu(x)\exp\left[bt - \frac{\mu(x)}{b}(e^{bt}-1)\right]$$

is the density function of the future lifetime of $(x)$ under Gompertz' law of mortality. We want to calculate the complete expectation of life, $\mathring{e}_x$, (defined in chapter 2) under Gompertz' law of mortality. Using (5) we have

$$\mathring{e}_x = \int_0^\infty {}_tp_x dt = \int_0^\infty \exp\left[-\frac{\mu(x)}{b}(e^{bt}-1)\right]dt$$

$$= \exp\left[\frac{\mu(x)}{b}\right]\int_0^\infty \exp\left[-\frac{\mu(x)}{b}e^{bt}\right]dt$$

In order to calculate this integral, we substitute $\frac{\mu(x)}{b}e^{bt}$ by $e^u$ and solve for $t$ to obtain

$$\begin{aligned} & \frac{\mu(x)}{b}e^{bt} = e^u \\ \Leftrightarrow\ & e^{bt} = e^u \cdot \frac{b}{\mu(x)} \\ \Leftrightarrow\ & bt = u + \ln\left[\frac{b}{\mu(x)}\right] \\ \Leftrightarrow\ & t = \frac{1}{b}\cdot u + \frac{1}{b}\ln\left[\frac{b}{\mu(x)}\right]. \end{aligned}$$

Thus, we have

$$\frac{dt}{du} = \frac{1}{b}.$$

Since we integrate with respect to t between the limits zero and $\infty$, the corresponding limits for $u = \ln[\frac{\mu(x)}{b}e^{bt}] = \ln[\frac{\mu(x)}{b}] + bt$ are $\ln[\frac{\mu(x)}{b}]$ and $\infty$ and we obtain

$$\overset{\circ}{e}_x = \exp\left[\frac{\mu(x)}{b}\right]\int_{\ln[\frac{\mu(x)}{b}]}^\infty \frac{1}{b}e^{-e^u}du = \frac{1}{b}e^{\frac{\mu(x)}{b}}H\left[\ln\left(\frac{\mu(x)}{b}\right)\right],$$

where $H(t) = \int_t^\infty e^{-e^u}du$.

**Theorem 2** *The function $H(\cdot)$ is strictly increasing and convex.*

**Proof:**

a) It is obvious that $H(\cdot)$ is strictly increasing, since $e^{-e^u}$ is positive.

b) To show that $H(\cdot)$ is strictly convex, we evaluate its first and second derivative:

$$H'(t) = -e^{-e^t}$$

and

$$H''(t) = -e^{-e^t} \cdot (-1) \cdot e^t = e^{-e^t} \cdot e^t > 0.$$

Since the second derivative of $H$ is strictly greater than zero, this completes the proof.

□

Let $\theta$ be equal to $\ln[\frac{\mu(x)}{b}]$, then

$$\overset{\circ}{e}_x = \frac{1}{b} H(\theta) e^{e^{\theta}} \tag{24}$$

In order to calculate the variance of the future lifetime, $Var[T(x)]$, using Gompertz law of mortality, we first have to find the second moment of $T(x)$: $E[T^2(x)]$. Using (6) we have

$$\begin{aligned} E[T^2(x)] = 2\int_0^{\infty} t \cdot {}_tp_x dt = 2\int_0^{\infty} t \exp\left[-\frac{\mu(x)}{b}(e^{bt}-1)\right] dt \\ = \quad 2e^{\mu(x)/b} \int_0^{\infty} t \exp\left[-\frac{\mu(x)}{b} e^{bt}\right] dt. \end{aligned}$$

Substituting $\frac{\mu(x)}{b} e^{bt}$ by $e^u$ and using that $\theta$ is equal to $\ln[\frac{\mu(x)}{b}]$ we get:

$$\begin{aligned} E[T^2(x)] &= 2e^{\mu(x)/b} \cdot \frac{1}{b^2} \cdot \int_{\ln[\frac{\mu(x)}{b}]}^{\infty} (u - \ln[\frac{\mu(x)}{b}]) \cdot \exp[-e^u] du \\ &= \frac{2}{b^2} \int_{\theta}^{\infty} (u-\theta) \exp[-e^u] du \cdot \exp[e^{\theta}] \qquad (25) \\ &= \frac{2}{b^2} G(\theta) \exp[e^{\theta}], \qquad (26) \end{aligned}$$

where $G(t) = \int_t^{\infty} (u-t) \exp[-e^u] du$

**Theorem 3** $G(\cdot)$ *is*

*(i) nonnegative,*

*(ii) strictly increasing,*

*(iii) convex, and*

*(iv)* $G'(t) = -H(t)$

**Proof:**

(i) Since $t \leq u < \infty$, this is obvious.

(ii) and (iv) Using (17) we have

$$\begin{aligned} G'(t) &= \int_t^\infty -\exp[-e^u]du + 0 - (t-t)\exp[-e^t] = -\int_t^\infty \exp[-e^u]du \\ &= -H(t) \end{aligned}$$

$\Rightarrow$ (iv). Since $H(t) > 0$, this implies that $G'(t) = -H(t) < 0$. Thus, $G(\cdot)$ is strictly decreasing.

(iii) Since

$$G''(t) \overset{(iv)}{=} -H'(t) = -(-exp[-e^t]) = exp[-e^t]$$

is greater than zero, $G(\cdot)$ is strictly convex.

$\square$

Using (24) and (26), we have

$$\begin{aligned} Var[T] = E(T^2) - E^2(T) = \frac{2}{b^2}G(\theta)\exp[e^\theta] - \overset{\circ}{e}{}_x^2 \\ = \overset{\circ}{e}{}_x^2\left[\frac{2G(\theta)}{\exp[e^\theta][H(\theta)]^2} - 1\right]. \end{aligned} \tag{27}$$

**Definition 5** *The variability coefficient* $\tau_T$ *is defined by*

$$\tau_T = \frac{\sqrt{Var(T)}}{\overset{\circ}{e}_x}. \tag{28}$$

The variability coefficient for the future lifetime $T$ is

$$\tau_T = \frac{\sqrt{\frac{2G(\theta)}{H^2(\theta)\exp[e^\theta]} - 1} \cdot \overset{\circ}{e}_x}{\overset{\circ}{e}_x} = \sqrt{\frac{2G(\theta)}{H^2(\theta)\exp[e^\theta]} - 1}. \tag{29}$$

**Theorem 4** $\lim_{\theta\to-\infty} \tau_T = 0$

**Proof:**

First, we note that

$$\lim_{\theta\to-\infty} H'(\theta) = \lim_{\theta\to-\infty} -\exp[-e^\theta] = -1.$$

Using the de l'Hospital Rule and Theorem 3 (iv) we have

$$\begin{aligned}\lim_{\theta\to-\infty} \frac{G(\theta)}{H^2(\theta)} &= \lim_{\theta\to-\infty} \frac{G'(\theta)}{2H(\theta)H'(\theta)} = \lim_{\theta\to-\infty} \frac{-H(\theta)}{2H(\theta)H'(\theta)} = \lim_{\theta\to-\infty} -\frac{1}{2H'(\theta)} \\ &= -\frac{1}{2(-1)} = \frac{1}{2}.\end{aligned}$$

Thus, $\tau_T = \sqrt{2 \cdot \frac{1}{2} \cdot 1 - 1} = 0$.

$\square$

This shows that under the assumption of Gompertz law and for small values of $\theta$, the variability coefficient $\tau_T$ is approximately 0.

We want to examine what happens for large values of $\theta$.

**Theorem 5** $\lim_{\theta\to\infty} \tau_T = 1$

**Proof:**

First, note that

$$\lim_{\theta\to\infty} H(\theta) = 0.$$

Using the de l'Hospital Rule we get

$$\begin{aligned}\lim_{\theta\to\infty}\frac{e^{-\theta}\exp\{-e^{\theta}\}}{H(\theta)} &= \lim_{\theta\to\infty}\frac{-e^{-\theta}\exp\{-e^{\theta}\}+\exp\{-e^{\theta}\}\cdot(-1)}{-\exp\{-e^{\theta}\}\cdot e6\theta\cdot e^{-\theta}}\\ &= \lim_{\theta\to\infty}[e^{-\theta}+1]=1.\end{aligned}$$

This implies that

$$\lim_{\theta\to\infty}\frac{e^{-\theta}H'(\theta)}{H(\theta)}=-1.$$

Thus, using the de l'Hospital Rule and Theorem 3 (iv), we have

$$\lim_{\theta\to\infty}\frac{G(\theta)}{e^{-\theta}H(\theta)}=\lim_{\theta\to\infty}\frac{G'(\theta)}{-e^{-\theta}H(\theta)+e^{-\theta}H'(\theta)}$$
$$=\lim_{\theta\to\infty}\frac{-H(\theta)}{-e^{-\theta}H(\theta)+e^{-\theta}H'(\theta)}=\lim_{\theta\to\infty}\frac{1}{e^{-\theta}-e^{-\theta}\frac{H'(\theta)}{H(\theta)}}=\frac{1}{0-(-1)}=1.$$

Hence,

$$\lim_{\theta\to\infty}\frac{G(\theta)\exp\{-e^{\theta}\}}{H^2(\theta)}=\lim_{\theta\to\infty}\frac{e^{-\theta}\exp\{-e^{\theta}\}}{H(\theta)}\cdot\frac{G(\theta)}{e^{-\theta}H(\theta)}=1\cdot1=1.$$

□

Finally,

$$\lim_{\theta\to\infty}\tau_T=\lim_{\theta\to\infty}\sqrt{\frac{2G(\theta)\exp\{-e^{\theta}\}}{H^2(\theta)}-1}=\sqrt{2\cdot1-1}=1.$$

This implies that $\frac{\sqrt{Var(T)}}{\mathring{e}_x}$ is approximately 1 the greater the value of $\theta$, which means the more $\mu(x)$ exceeds b.

3.2 Multiple Lives

Again we assume that mortality follows Gompertz's law. Besides we assume that $T(x)$ and $T(y)$ are independent. We have already seen in equation (21) that the force of mortality is equal to $Bc^x$ if mortality follows Gompertz. We want to substitute a joint-life status $(xy)$ by a single-life survival status $(w)$ that has a force of mortality equal to the force of mortality of $(xy)$ for all $t \geq 0$. Consider $\mu_{xy}(s) = \mu(w+s)$, $s \geq 0$. Since $T(x)$ and $T(y)$ are assumed to be independent, we know from Theorem 1 that this equation is equivalent to:

$$\begin{aligned} \mu(x+s) + \mu(y+s) = \mu(w+s) \quad &\Leftrightarrow \quad Bc^{x+s} + Bc^{y+s} = Bc^{w+s} \\ &\Leftrightarrow \quad c^x + c^y = c^w \end{aligned} \tag{30}$$

(30) defines w. It follows that for $t \geq 0$,

$$\begin{aligned} {}_tp_w &= \exp[-\int_0^t \mu(w+s)ds] \\ &= \exp[-\int_0^t \mu_{xy}(s)ds] \\ &= {}_tp_{xy}. \end{aligned}$$

This implies that if $w$ is defined as in (30), then all probabilities, expected values, and variances for the joint-life status $(xy)$ equal those for the single life $(w)$. Let's look at an example:

**Example 1** *Calculate the value of* $\ddot{a}_{60:70}$ *if the interest rate is 6 percent and if* $c = 10^{0.04}$ *using Gompertz's law of mortality and assuming independence of the future lifetimes T(x) and T(y).*

**Solution:**

$$(30) \Rightarrow (10^{0.04})^{60} + (10^{0.04})^{70} = (10^{0.04})^{w}$$

*This simplifies to:*

$$882.146 = (10^{0.04})^{w} \Leftrightarrow w = \frac{\ln[882.146]}{\ln[10^{0.04}]} = 73.63851158.$$

*Using linear interpolation, we have:*

$$\ddot{a}_{60:70} = \sum_{k=0}^{\infty} v^k \cdot {}_kp_{60:70} = 0.63851158\ddot{a}_{74} + 0.36148842\ddot{a}_{73} = 7.584.$$

*The value by the* $\ddot{\mathrm{a}}_{60}$ *table is* 7.55633*.*

# CHAPTER IV

# LIKELIHOOD RATIO TESTS AND OTHER STATISTICAL TESTS

In this section we want to present two kinds of tests: Likelihood-ratio tests as in Mood, Graybill and Boes (1974) and a distribution-free test for independence as in Hollander and Wolfe (1999).

## 4.1 Likelihood-Ratio Tests

Let $\vartheta$ be an unknown parameter vector and let $\theta$, called the *parameter space*, denote the set of possible values that $\vartheta$ can assume.

**Definition 6** *The likelihood function of n random variables $X_1, ..., X_n$ is defined to be the joint density of the n random variables, denoted by*

$$L(\vartheta) = f_{X_1,...,X_n}(x_1, ..., x_n; \vartheta),$$

*where $\vartheta$ is the unknown parameter vector.*

**Definition 7** *Let $L(\vartheta) = L(\vartheta; x_1, ..., x_n)$ be the likelihood function for the random variables $X_1, ..., X_n$. Let $\hat{\vartheta} = \hat{\vartheta}(x_1, ..., x_n)$ denote the value of $\vartheta$ in $\theta$ that maximizes $L(\vartheta)$, i.e. $L(\hat{\vartheta}) \geq L(\vartheta)$ for all $\vartheta \in \theta$. Then $T = \hat{\vartheta}(X_1, ..., X_n)$ is the maximum likelihood estimator of $\vartheta$. $\hat{\vartheta} = \hat{\vartheta}(x_1, ..., x_n)$ is the maximum likelihood estimate for $\vartheta$ for the sample $x_1, ..., x_n$.*

**Remark 2** *The maximum likelihood estimate $\hat{\vartheta}(x_1, ..., x_n)$ is the value of $\theta$ that is* *"most likely" to have produced the data set $x_1, ..., x_n$.*

It is sometimes easier to find the maximum of the logarithm of the likelihood function, denoted by $\ln[L(\vartheta)] = l(\vartheta)$, rather than working with the likelihood function itself. Since the logarithm is a monotonic function, $L(\vartheta)$ and $l(\vartheta)$ have the same maxima. Before we look at an example, we need another definition.

**Definition 8** *A set of $n$ independent and identically distributed random variables is called a random sample (Everitt, 2002).*

Let's look at an example.

**Example 2** *Suppose that we draw a random sample of size $n$ from the normal distribution with mean $\mu$ and variance of 1. $\mu$ is the only unknown parameter. The sample values are denoted by $x_1, ..., x_n$. Since we have a random sample, the likelihood function is the product of $n$ normal density functions:*

$$\begin{aligned} L(\mu) &= \prod_{i=1}^{n}(2\pi)^{-\frac{1}{2}} \exp\left[-\frac{(x_i-\mu)^2}{2}\right] \\ &= (2\pi)^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right] \\ l(\mu) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2 \end{aligned}$$

*The necessary condition for a maximum is that the derivative of $l(\mu)$ with respect to $\mu$ has to be equal to 0.*

$$\begin{aligned} \frac{d}{d\mu}l(\mu) = 0 &\Leftrightarrow \sum_{i=1}^{n}(x_i-\hat{\mu}) = 0 \\ &\Leftrightarrow n\cdot\frac{1}{n}\sum_{i=1}^{n}x_i - n\hat{\mu} = 0 \\ &\Leftrightarrow n\bar{x} - n\hat{\mu} = 0 \Leftrightarrow \hat{\mu} = \bar{x}. \end{aligned}$$

*We still have to check if the sufficient condition for a maximum holds:*

$$\frac{d^2}{d^2\mu} l(\hat{\mu}) = -n < 0$$

*This implies that $\hat{\mu} = \bar{x}$ is the maximum likelihood estimate.*

4.1.1 Generalized Likelihood Ratio Tests

The generalized likelihood ratio test tests composite hypotheses. Let $X_1, ..., X_n$ be a sample with joint density $f_{X_1,...,X_n}(x_1, ..., x_n; \vartheta)$. Let $\vartheta \in \theta$. Suppose that $\theta$ is the $k$-dimensional parameter space. Suppose we want to test $\{H_0 : \vartheta_1 = \vartheta_1^0, ..., \vartheta_r = \vartheta_r^0, \vartheta_{r+1}, ..., \vartheta_k\}$, where $\vartheta_1^0, ..., \vartheta_r^0$ are known and $\vartheta_{r+1}, ..., \vartheta_k$ are left unspecified. Let $\theta_0 = \{\vartheta : \vartheta \in H_0\}$.

**Definition 9** *Let $L(\vartheta; x_1, ..., x_n)$ be the likelihood function for the random variables $X_1, ..., X_n$. Let $f_{X_1,...,X_n}(x_1, ..., x_n; \vartheta)$ denote their joint density function. The generalized likelihood-ratio, denoted by $\lambda$, is defined by (Mood et al. 1974)*

$$\lambda = \lambda(x_1, ..., x_n) = \frac{sup\{L(\vartheta; x_1, ..., x_n) : \vartheta \in \theta_0\}}{sup\{L(\vartheta; x_1, ..., x_n) : \vartheta \in \theta\}}.$$

We replace the observations $x_1, ..., x_n$ by the random variables $X_1, ..., X_n$, we write $\Lambda = \lambda(X_1, ..., X_n)$.

**Remark 3** *(i) $0 \leq \lambda \leq 1$;*

*$\lambda \geq 0$: because both numerator and denominator are nonnegative*

*$\lambda \leq 1$: the supremum taken in the numerator is over a smaller set of parameter values than the supremum in the denominator.*

*(ii) The denominator of $\lambda$ is the likelihood function evaluated at the maximum likelihood estimate $\hat{\vartheta}$. If the numerator is much smaller than the denominator, which means that $\lambda$ is small, the data $x_1, ..., x_n$ do not support the null hypothesis. So*

*$H_0$ should be rejected whenever $\lambda \leq \lambda_0$, where $\lambda_0$ is some fixed constant satisfying $0 \leq \lambda_0 \leq 1$.*

Asymptotic distribution of generalized likelihood ratio

$$-2\log[\Lambda] = -2\log\left[\frac{sup\{L(\vartheta; X_1, ..., X_n) : \vartheta \in \theta_0\}}{sup\{L(\vartheta; X_1, ..., X_n) : \vartheta \in \theta\}}\right]$$

is asymptotically distributed as a chi-square distribution with $r$ degrees of freedom, when $H_0$ is true and the sample size $n$ is large (Mood et al., 1974). The degrees of freedom, $r$, can be interpreted in two ways: (i) as the number of parameters specified by $H_0$ and (ii) as the difference in the dimensions of $\theta$ and $\theta_0$.

In Remark 3 (ii) $H_0$ is rejected for small values of $\lambda$. Since $-2\log[\lambda]$ decreases in $\lambda$ a test that is equivalent to a generalized likelihood test is one that rejects $H_0$ for large values of $-2\log[\lambda]$. Thus, a test with approximate significance level $\alpha$ is given by the following: Reject $H_0$ whenever $-2\log[\lambda] > \chi^2_\alpha(r)$, where $\chi^2_\alpha(r)$ is the $\alpha$-quantile of the chi-squared distribution with $r$ degrees of freedom. Let's look at an example:

**Example 3** *Suppose that a random sample of size $n$ is drawn from a normal distribution like in Example 2. We want to test $\{H_0 : \mu = 0$ versus $H_1 : \mu \neq 0\}$. The maximum likelihood estimate is $\hat{\mu} = \bar{x}$. The test statistic for the asymptotic generalized likelihood ratio-test is*

$$\begin{aligned}
& -2\ln\left[\frac{L(0; x_1, ..., x_n)}{L(\hat{\mu}; x_1, ..., x_n)}\right] \\
= \quad & -2\ln\left[\frac{(2\pi)^{-\frac{n}{2}}\exp\left[-\sum_{i=1}^{n}(x_i - 0)^2/2\right]}{(2\pi)^{-\frac{n}{2}}\exp\left[-\sum_{i=1}^{n}(x_i - \bar{x})^2/2\right]}\right]
\end{aligned}$$

$$\begin{aligned}
&= -2\ln\left[\exp[-\sum_{i=1}^{n}(x_i^2)/2]\right] + 2\ln\left[\exp[-\sum_{i=1}^{n}(x_i-\bar{x})^2/2]\right] \\
&= \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n}(x_i-\bar{x})^2 \\
&= \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\
&= 2\sum_{i=1}^{n} x_i\bar{x} - n\bar{x}^2 = 2n\bar{x}^2 - n\bar{x}^2 = n\bar{x}^2.
\end{aligned}$$

*This test statistic has a chi-squared distribution with one degree of freedom because one parameter ($\mu$) was specified in the reduced model. Reject $H_0$ whenever*

$$-2\log[\lambda] = n\bar{x}^2 > \chi^2_\alpha(1).$$

*If we choose $\alpha$ to be 0.05, we get: Reject $H_0$ whenever*

$$n\bar{x}^2 > 3.84. \tag{31}$$

4.2 A Distribution-Free Test for Independence

Let $(X_1, Y_1), ..., (X_n, Y_n)$ be a random sample from a continuous bivariate population with joint distribution $F_{XY}$ and marginal distributions $F_X$ and $F_Y$. The null hypothesis is

$$H_0 : F_{XY}(x, y) = F_X(x)F_Y(y) \text{ for all } (x,y) \text{ pairs.} \tag{32}$$

We introduce Kendall's Tau and Spearman's rho because we are using these correlation coefficients for measuring the dependence of the future lifetimes of a couple in section 5.6.

4.2.1 Kendall's Tau

The alternative will be that type of dependence between $X$ and $Y$ which is of principal interest. In this section, we concentrate on a type of dependence measured by Kendall's correlation coefficient.

**Definition 10** *Kendall's correlation coefficient is defined by*

$$\tau = 2P[(Y_2 - Y_1)(X_2 - X_1) > 0] - 1, \tag{33}$$

*where* $(X_1, Y_1), ..., (X_n, Y_n)$ *is a random sample from a continuous bivariate population.*

**Theorem 6** *If $X$ and $Y$ are independent, then $\tau$ is equal to 0.*

**Proof:**

The event $\{(Y_2 - Y_1)(X_2 - X_1) > 0\}$ occurs if and only if $\{Y_2 > Y_1 \,\text{and}\, X_2 > X_1\}$ or $\{Y_2 < Y_1 \,\text{and}\, X_2 < X_1\}$. Since these events are mutually exclusive, we have:

$$P[(Y_2 - Y_1)(X_2 - X_1) > 0] = P[X_2 > X_1, Y_2 > Y_1] + P[X_2 < X_1, Y_2 < Y_1]$$

If $X$ and $Y$ are independent, we have

$$P[X_2 > X_1, Y_2 > Y_1] = P[X_2 > X_1]P[Y_2 > Y_1] = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4},$$

because $X_1$ and $X_2$ are independent and identically distributed (iid) and $Y_1, Y_2$ are iid as well. Note that $Y_1, Y_2$ need not have the same distribution as $X_1$ and $X_2$. Similarly, if $X$ and $Y$ are independent, it follows that

$$P[X_2 < X_1, Y_2 < Y_1] = \frac{1}{4}.$$

Thus, if $X$ and $Y$ are independent, we have that

$$\tau = 2\left(\frac{1}{4} + \frac{1}{4}\right) - 1 = 0.$$

$\square$

**Definition 11** *The Kendall statistic $K$ is defined by (Hollander and Wolfe, 1999)*

$$K = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Q[(X_i, Y_i), (X_j, Y_j)], \tag{34}$$

*where*

$$Q[(a,b),(c,d)] = \begin{cases} 1 & , \textit{ if (d-b)(c-a)} > 0 \\ -1 & , \textit{ if (d-b)(c-a)} < 0 \end{cases} \tag{35}$$

This means that for each pair of subscripts $(i, j)$ with $i < j$, score 1 if $(Y_j - Y_i)(X_j - X_i)$ is positive and score $-1$ if it is negative. Thus, $K$ adds up the 1s and $-1$s from the paired sign statistics. There are three possible types of tests:

a. One-sided upper-tail test: We want to test (32), which implies $\tau = 0$ versus the alternative that $X$ and $Y$ are positively correlated, i.e. $H_1 : \tau > 0$. We reject $H_0$ whenever $K \geq k_\alpha$ at the level of significance $\alpha$, where $k_\alpha$ is chosen such that the type 1 error probability is equal to $\alpha$. To motivate this test: The null hypothesis is that the $X$ and $Y$ random variables are independent, which implies that $\tau$ is equal to zero. The alternative in this procedure is that $\tau$ is positive, which implies that

$P[(Y_2 - Y_1)(X_2 - X_1) > 0] > \frac{1}{2}$. Thus, there tend to be a large number of positive paired sign statistics and fewer negative paired sign statistics. Hence, we expect a big, positive value for $K$. This suggests that we should reject $H_0$ in favor of $\tau > 0$ for large values of $K$.

b. One-sided lower-tail test: We want to test (32) versus the alternative that $X$ and $Y$ are negatively correlated, i.e. $H_2 : \tau < 0$.
We reject $H_0$ whenever $K \leq -k_\alpha$ at the level of significance $\alpha$.

c. Two-sided test: We want to test (32) versus the alternative that $X$ and $Y$ are dependent, i.e. $H_3 : \tau \neq 0$. We reject $H_0$ whenever $|K| \geq k_{\alpha/2}$ at the level of significance $\alpha$.

To justify test procedures b. and c. note that the distribution of $K$ under the null hypothesis is symmetric about 0 (Hollander and Wolfe, 1999). This implies

$$P(K \leq -x) = P(K \geq x)$$

under the null hypothesis. Thus, we have

$$\begin{aligned} P(|K| \geq k_{\alpha/2}) = 1 - P(-k_{\alpha/2} \leq K \leq k_{\alpha/2}) &= 1 - [P(K \leq k_{\alpha/2}) - P(K \leq -k_{\alpha/2})] \\ &= 1 - [1 - \alpha/2 - \alpha/2] = \alpha. \end{aligned}$$

Another possibility of testing those three hypotheses is using a large sample approximation. It is based on the asymptotic normality of $K$. To standardize $K$, we need to know the expected value and variance of $K$ when the null hypothesis of independence is true. The expected value and variance are given in Hollander and Wolfe (1999):

**Lemma 4** *Under $H_0$, the expected value and variance of $K$ are*

$$(i)\, E_0(K) = 0$$

*and*

$$(ii)\, Var_0(K) = \frac{n(n-1)(2n+5)}{18}.$$

**Proof:**

$$\begin{aligned}
(i) \qquad E(K) &= E\{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Q[(X_i,Y_i),(X_j,Y_j)]\} \\
&= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} E\{Q[(X_i,Y_i),(X_j,Y_j)]\} \\
&= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \{P[(Y_2-Y_1)(X_2-X_1)>0] - P[(Y_2-Y_1)(X_2-X_1)<0]\} \\
&= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \{P[(Y_2-Y_1)(X_2-X_1)>0] \\
&- P[(Y_2-Y_1)(X_2-X_1)<0] + 1 - 1\} \\
&= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \{P[(Y_2-Y_1)(X_2-X_1)>0] - P[(Y_2-Y_1)(X_2-X_1)<0] \\
&+ P[(Y_2-Y_1)(X_2-X_1)>0] + P[(Y_2-Y_1)(X_2-X_1)<0] - 1\} \\
&= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \{2P[(Y_2-Y_1)(X_2-X_1)>0] - 1\} \\
&= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \tau = \sum_{i=1}^{n-1}(n-i)\tau = \sum_{j=1}^{n-1} j\tau = \frac{(n-1)n}{2}\tau = \binom{n}{2}\tau
\end{aligned}$$

Under the null hypothesis of independence we have

$$E_0[K] = 0$$

$$(ii) \qquad Var(K) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Var(Q_{ij}) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{s=1}^{n-1} \sum_{t=s+1}^{n} Cov(Q_{ij}, Q_{st}),$$
$$\text{for } (i,j) \neq (s,t),$$

where

$$Q_{uv} = Q[(X_u, Y_u), (X_v, Y_v)], \text{ for } 1 \leq u < v \leq n.$$

The following equality holds (Hollander and Wolfe, 1999):

$$Var(K) = [n(n-1)\left[\frac{1}{2}(1-\tau^2) + 4(n-2)\left\{\delta - \left(\frac{\tau+1}{2}\right)^2\right\}\right],$$
$$\text{where}\, \delta = P[(Y_2 - Y_1)(X_2 - X_1) > 0\, \text{and}\, (Y_3 - Y_1)(X_3 - X_1) > 0]$$

We can break down $\delta$:

$$\begin{aligned}
\delta &= Pr[Y_2 > Y_1, X_2 > X_1, Y_3 > Y_1, X_3 > X_1] \\
&+ Pr[Y_2 > Y_1, X_2 > X_1, Y_3 < Y_1, X_3 < X_1] \\
&+ Pr[Y_2 < Y_1, X_2 < X_1, Y_3 > Y_1, X_3 > X_1] \\
&+ Pr[Y_2 < Y_1, X_2 < X_1, Y_3 < Y_1, X_3 < X_1] \\
&= Pr[Y_1 < \min(Y_2, Y_3), X_1 < \min(X_2, X_3)] \\
&+ Pr[Y_2 > Y_1 > Y_3, X_2 > X_1 > X_3] \\
&+ Pr[Y_2 < Y_1 < Y_3, X_2 < X_1 < X_3] \\
&+ Pr[Y_1 > \max(Y_2, Y_3), X_1 > \max(X_2, X_3)]
\end{aligned}$$

When $X$ and $Y$ are independent, we have:

$$\begin{aligned}
\delta_0 &= Pr_0[Y_1 < \min(Y_2, Y_3)]Pr_0[X_1 < \min(X_2, X_3)] \\
&+ Pr_0[Y_2 > Y_1 > Y_3]Pr_0[X_2 > X_1 > X_3] \\
&+ Pr_0[Y_2 < Y_1 < Y_3]Pr_0[X_2 < X_1 < X_3] \\
&+ Pr_0[Y_1 > \max(Y_2, Y_3)]Pr_0[X_1 > \max(X_2, X_3)].
\end{aligned}$$

Since $X_1, X_2, X_3$ are mutually independent and identically distributed, as are $Y_1, Y_2, Y_3$, we have:

$$Pr_0[Y_1 < \min(Y_2, Y_3)] = Pr_0[X_1 < \min(X_2, X_3)] = \frac{1}{3},$$
$$Pr_0[Y_1 > \max(Y_2, Y_3)] = Pr_0[X_1 > \max(X_2, X_3)] = \frac{1}{3},$$

and

$$\begin{aligned}
Pr_0[Y_2 > Y_1 > Y_3] = Pr_0[X_2 > X_1 > X_3] = Pr_0[Y_2 < Y_1 < Y_3] \\
= Pr_0[X_2 < X_1 < X_3] = \frac{1}{6}.
\end{aligned}$$

This implies

$$\delta_0 = \frac{1}{3}\left(\frac{1}{3}\right) + \frac{1}{6}\left(\frac{1}{6}\right) + \frac{1}{6}\left(\frac{1}{6}\right) + \frac{1}{3}\left(\frac{1}{3}\right) = \frac{10}{36}.$$

Since we assume that $X$ and $Y$ are independent we substitute $\tau$ by zero and $\delta_0$ by $\frac{10}{36}$.

$$\begin{aligned} Var_0(K) &= n(n-1)\left[\frac{1}{2}(1-0^2)+4(n-2)\left\{\frac{10}{36}-\left(\frac{0+1}{2}\right)^2\right\}\right] \\ &= n(n-1)\left[\frac{1}{2}+\frac{1}{9}(n-2)\right] = \frac{n(n-1)(2n+5)}{18} \end{aligned}$$

□

Hence, the standardized version of $K$ is

$$K^* = \frac{K - E_0(K)}{\sqrt{Var_0(K)}} = \frac{K}{\sqrt{\frac{n(n-1)(2n+5)}{18}}} \tag{36}$$

When $H_0$ is true and $n$ tends to infinity, then $K^*$ has an asymptotic $N(0,1)$ distribution (Hollander and Wolfe, 1999). That's why for a large sample size, our test looks like that:

a. Reject $H_0$ in favor of the alternative that $X$ and $Y$ are positively correlated whenever $K^* \geq z_\alpha$, where $z_\alpha$ is the $\alpha$-quantile of $N(0,1)$.

b. Reject $H_0$ in favor of the alternative that $X$ and $Y$ are negatively correlated whenever $K^* \leq -z_\alpha$.

c. Reject $H_0$ in favor of the alternative that $X$ and $Y$ are dependent whenever $|K^*| \geq z_{\alpha/2}$.

There is still the question remaining: What happens to $K$ if we have ties among the $n$ $X$ observations and/or among the $n$ $Y$ observations? We replace the function $Q[(a,b),(c,d)]$ in the definition of $K$ (34) by

$$Q^*[(a,b)(c,d)] = \begin{cases} 1 & \text{, if (d-b)(c-a)} > 0 \\ 0 & \text{, if (d-b)(c-a)} = 0 \\ -1 & \text{, if (d-b)(c-a)} < 0 \end{cases} \tag{37}$$

This means that in the case of tied $X$ or $Y$ values, zeros are assigned to the associated paired sign statistics. When we now use procedure a., b. or c. and when we substitute $Q$ by $Q^*$ in Definition 11, this test has only approximately a level of significance of $\alpha$ because the ties affect the variance under the null hypothesis of independence of $K$ (Hollander and Wolfe, 1999). An estimator of Kendall's correlation coefficient $\tau$ is (Hollander and Wolfe, 1999)

$$\hat{\tau} = \frac{2K}{n(n-1)}. \tag{38}$$

**Lemma 5** *$\hat{\tau}$ assumes only values between -1 and 1.*

**Proof:**

Since $K = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Q[(X_i, Y_i), (Y_i, Y_j)]$ we get the smallest value of $K$ when $Q[(X_i, Y_i), (Y_i, Y_j)]$ is -1 for all $i, j$. Then we have $K = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} -1 = -\frac{n(n-1)}{2}$. This implies that $\hat{\tau} = \frac{2K}{n(n-1)} = \frac{-\frac{n(n-1)}{2} \cdot 2}{n(n-1)} = -1$. The argument is similar for the maximum value of $\hat{\tau}$.

□

If there are ties, we should use $Q^*[(a, b), (c, d)]$ as in (37) in the definition of $K$.

4.2.2 Spearman's Rho

Another possibility of testing independence is using Spearman's rho.

**Definition 12** *Spearman's correlation coefficient (rho) is defined by*

$$\rho_s = 6Pr[(X_1 - X_2)(Y_1 - Y_3) > 0] - 3, \tag{39}$$

*where $(X_1, Y_1), ..., (X_n, Y_n)$ is a random sample from a continuous bivariate population.*

**Theorem 7** *If $X$ and $Y$ are independent, Spearman's rho is zero.*

**Proof:**

We know from the proof of Theorem 6 that if $X$ and $Y$ are independent, we have

$$\begin{aligned} & Pr[(X_1 - X_2)(Y_1 - Y_3) > 0] \\ = \; & Pr[X_2 > X_1, Y_2 > Y_1] + Pr[X_2 < X_1, Y_2 < Y_1] = \frac{1}{2}. \end{aligned}$$

Thus, we have

$$\rho_s = 6Pr[(X_1 - X_2)(Y_1 - Y_3) > 0] - 3 = 6\left(\frac{1}{2}\right) - 3 = 0$$

□

We present the classical estimate of Spearman's rho and the proof (following Kruskal, 1958).

**Lemma 6** *An estimate of Spearman's rho*

$$\rho_s = 6Pr[(X_1 - X_2)(Y_1 - Y_3)] - 3, \tag{40}$$

*is*

$$\hat{\rho}_s = \frac{\sum_{k=1}^{n}[R_{X_i} - \frac{n+1}{2}][R_{Y_i} - \frac{n+1}{2}]}{(n^2 - 1)n/12}, \tag{41}$$

*where $(X_1, Y_1), ..., (X_n, Y_n)$ is a random sample from a continuous bivariate population, $R_{X_i}$ is the rank of $X_i$ and $R_{Y_i}$ is the rank of $Y_i$.*

**Proof**:

We want to estimate Spearman's correlation coefficient. At first we should find an estimate for

$$Pr[(X_1 - X_2)(Y_1 - Y_3) > 0] = Pr[(X_1, X_2)(Y_1, Y_3)\,\text{concordant}],$$

where two hypothetical bivariate observations $(X_1, X_2)(Y_1, Y_3)$ are *concordant*, in the sense that the two $x$-coordinates differ with the same sign as the two $y$-coordinates. Kruskal (1958) considers actual observations $(X_i, Y_i)$. There are $n^2 - 1$ points of the form $(X_j, Y_k)$ excluding the point $(X_i, Y_i)$, we work with. Replace the $X_i's$ by the numbers $1, 2, ..., n$ (lowest $X_i$ is replaced by 1, next lowest by 2, and so on). These ranks of $X_i$ are denoted by $R_{X_i}$. Similarly, the $Y_i's$ are replaced by their ranks. Denote these ranks by $R_{Y_i}$. From these $n^2 - 1$ points $(X_j, Y_k)$, exactly $(R_{X_i} - 1)(R_{Y_i} - 1)$ points will lie below and to the left of $(X_i, Y_i)$ and $(n - R_{X_i})(n - R_{Y_i})$ points will lie above and to the right of $(X_i, Y_i)$ . Let's look at an example for $n = 4$:

**Example 4** *Imagine we have 4 x- and 4 y-coordinates:* $X_1$*=1.2,* $X_2$*=2.3,* $X_3$*=4.5,* $X_4$*=2.2,* $Y_1$*=2,* $Y_2$*=4.3,* $Y_3$*=2.4,* $Y_4$*=1.4. Thus, we have 16 points:* $(X_1, Y_1) = (1.2, 2)$, $(X_1, Y_2) = (1.2, 4.3)$, $(X_1, Y_3) = (1.2, 2.4)$, $(X_1, Y_4) = (1.2, 1.4)$, $(X_2, Y_1) = (2.3, 2)$, $(X_2, Y_2) = (2.3, 4.3)$, $(X_2, Y_3) = (2.3, 2.4)$, $(X_2, Y_4) = (2.3, 1.4)$, $(X_3, Y_1) = (4.5, 2)$, $(X_3, Y_2) = (4.5, 4.3)$, $(X_3, Y_3) = (4.5, 2.4)$, $(X_3, Y_4) = (4.5, 1.4)$, $(X_4, Y_1) = (2.2, 2)$, $(X_4, Y_2) = (2.2, 4.3)$, $(X_4, Y_3) = (2.2, 2.4)$, $(X_4, Y_4) = (2.2, 1.4)$.

*Consider the point* $(X_2, Y_3) = (2.3, 2.4)$. *Since* $X_2$ *is the third smallest* $x$*-coordinate and* $Y_3$ *is the third smallest* $y$*-coordinate, we have* $R_{X_2} = 3$ *and* $R_{Y_3} = 3$. *There are*

$$(3 - 1) \cdot (3 - 1) = 4$$

*points where the* $x$*-coordinate is less than* $X_2$ *and the* $y$*-coordinate is less than* $Y_3$*;*

*namely* $(X_1, Y_1)$, $(X_1, Y_4)$ $(X_4, Y_1)$ $(X_4, Y_4)$. *There is*

$$(4-3) \cdot (4-3) = 1$$

*point where the* $x$*-coordinate is greater than* $X_2$ *and the* $y$*-coordinate is greater than* $Y_3$*; namely* $(X_3, Y_2)$. *Therefore, the number of points concordant with* $(X_2, Y_3)$ *is:*

$$4 + 1 = 5.$$

*There are still*

$$2(4-1) = 6$$

*points unequal to* $(X_2, Y_3)$, *where the* $x$*-coordinate is equal to* $X_2$ *or the* $y$*-coordinate is equal to* $Y_3$. *They lie between concordance and disconcordance.*

In general, we have that the number of points concordant with $(X_i, Y_i)$ is

$$\begin{aligned} & (R_{X_i} - 1)(R_{Y_i} - 1) + (n - R_{X_i})(n - R_{Y_i}) \\ = \; & 2R_{X_i}R_{Y_i} - (n+1)(R_{X_i} + R_{Y_i}) + n^2 + 1. \end{aligned}$$

There are still $2(n-1)$ points $(X_i, Y_i)$ between concordance and disconcordance. Since they lie between concordance and disconcordance, Kruskal (1958) counts them half. Therefore an estimate of $Pr[(X_1, X_2)(Y_1, Y_3)\,\text{concordant}]$ is the total number of intersections concordant with $(X_i, Y_i)$ added by one-half times the number of points between concordance and disconcordance divided by the total number of intersections excluding $(X_i, Y_i)$. Expressed as a formula, we get

$$\begin{aligned}
& \frac{1}{n^2-1}[2R_{X_i}R_{Y_i} - (n+1)(R_{X_i}+R_{Y_i}) + n^2 + 1 + \frac{1}{2}(2(n-1))] \\
= \;& \frac{1}{n^2-1}[2R_{X_i}R_{Y_i} - (n+1)(R_{X_i}+R_{Y_i}) + n^2 + 1 + n - 1] \\
= \;& \frac{1}{n^2-1}[2R_{X_i}R_{Y_i} - (n+1)(R_{X_i}+R_{Y_i}) + n(n+1)].
\end{aligned}$$

We still have to average this over the n $(X_i, Y_i)'s$ which can be considered:

$$\begin{aligned}
& \frac{1}{n(n^2-1)}\sum_{i=1}^{n}[2R_{X_i}R_{Y_i} - (n+1)(R_{X_i}+R_{Y_i}) + n(n+1)] \\
= \;& \frac{1}{n(n^2-1)}[2\sum_{i=1}^{n} R_{X_i}R_{Y_i} - (n+1)\sum_{i=1}^{n}(R_{X_i}+R_{Y_i}) + n^2(n+1)] \\
= \;& \frac{1}{n(n^2-1)}[2\sum_{i=1}^{n} R_{X_i}R_{Y_i} - (n+1)\sum_{i=1}^{n}(i+i) + n^2(n+1)] \\
= \;& \frac{1}{n(n^2-1)}[2\sum_{i=1}^{n} R_{X_i}R_{Y_i} - (n+1)2\frac{n(n+1)}{2} + n^2(n+1)] \\
= \;& \frac{1}{n(n^2-1)}[2\sum_{i=1}^{n} R_{X_i}R_{Y_i} - n(n+1)^2 + n^2(n+1)] \\
= \;& \frac{1}{n(n^2-1)}[2\sum_{i=1}^{n} R_{X_i}R_{Y_i} - n(n+1)(n+1-n)] \\
= \;& \frac{1}{n(n^2-1)}[2\sum_{i=1}^{n} R_{X_i}R_{Y_i} - n(n+1)]
\end{aligned}$$

To get the estimate of $\rho_s$, we can now use the relationship

$$\rho_s = 6Pr[(X_1 - X_2)(Y_1 - Y_3) > 0] - 3.$$

An estimate of $\rho_s$ is

$$
\begin{aligned}
\hat{\rho}_s &= 6[\frac{1}{n(n^2-1)}[2\sum_{i=1}^{n} R_{X_i}R_{Y_i} - n(n+1)]] - 3 \\
&= \frac{12\sum_{i=1}^{n} R_{X_i}R_{Y_i} - (6n(n+1) + 3n(n^2-1))}{n(n^2-1)} \\
&= \frac{\sum_{i=1}^{n} R_{X_i}R_{Y_i} - \frac{1}{4}n(n+1)(1+n)}{n(n^2-1)/12} \\
&= \frac{\sum_{i=1}^{n} R_{X_i}R_{Y_i} - \frac{n(n+1)^2}{4}}{n(n^2-1)/12} \\
&= \frac{\sum_{i=1}^{n} R_{X_i}R_{Y_i} - 2\frac{n(n+1)^2}{4} + \frac{n(n+1)^2}{4}}{n(n^2-1)/12} \\
&= \frac{\sum_{i=1}^{n} R_{X_i}R_{Y_i} - \frac{n+1}{2}\frac{n(n+1)}{2} - \frac{n+1}{2}\frac{n(n+1)}{2} + [\frac{n+1}{2}]^2 n}{n(n^2-1)/12} \\
&= \frac{\sum_{i=1}^{n} R_{X_i}R_{Y_i} - \frac{n+1}{2}\sum_{i=1}^{n} R_{X_i} - \frac{n+1}{2}\sum_{i=1}^{n} R_{Y_i} + [\frac{n+1}{2}]^2 n}{n(n^2-1)/12} \\
&= \frac{\sum_{i=1}^{n}[R_{X_i} - \frac{n+1}{2}][R_{Y_i} - \frac{n+1}{2}]}{(n^2-1)n/12}.
\end{aligned}
$$

□

The $n$ $X$ observations are ordered from least to greatest and $R_{X_i}$ is the rank of $X_i$, $i = 1, ..., n$, in this ordering. Similarly, the $n$ $Y$ observations are ordered and $R_{Y_i}$ denotes the rank of $Y_i$. If we reject $\{H_0$: $X$ and $Y$ are independent$\}$ in favor of $\{H_1$: $X$ and $Y$ are not independent$\}$ whenever $|\hat{\rho}_s| \geq r_{s,\frac{\alpha}{2}}$, this test has a level of significance of $\alpha$, where $r_{s,\frac{\alpha}{2}}$ is chosen to make the type 1 error probability equal to $\alpha$. Motivation of this test: Our null hypothesis is that $X$ and $Y$ are independent. This implies that any permutation of the $X$ ranks $(R_{X_1}, ..., R_{X_n})$ is equally likely to occur with any permutation of the Y ranks $(R_{Y_1}, ..., R_{Y_n})$. This means that $\hat{\rho}_s$ should be close to zero. In contrast to that, when the alternative: $\{X$ and $Y$ are not independent$\}$ is true, $|\hat{\rho}_s|$ should be large.

Similarly to the tests using Kendall's Tau, another possibility of testing is with a large sample approximation. It is based on the asymptotic normality of $\hat{\rho}_s$. To standardize $\hat{\rho}_s$ we need to know its expected value and variance when the null hypothesis of independence is true.

**Lemma 7** *Under $H_0$, the expected value and variance of $\hat{\rho}_s$ are:*

$$E(\hat{\rho}_s) = 0$$

*and*

$$Var(\hat{\rho}_s) = \frac{1}{n-1}.$$

So, this means that $\frac{\hat{\rho}_s - 0}{\sqrt{\frac{1}{n-1}}}$ is approximately standard normal distributed as $n$ tends to infinity. If we reject $H_0$ whenever $|\hat{\rho}_s \cdot \sqrt{n-1}| \geq z_{\alpha/2}$, this test has approximately a level of significance of $\alpha$ (Hollander and Wolfe, 1999).

# CHAPTER V

# BIVARIATE MORTALITY MODEL FOR COUPLED LIVES

## 5.1 Introduction and Notation

The following chapter discusses several ideas of Carriere (1986, 2000) and Frees et al. (1996) about modelling the dependence of the future lifetime of coupled lives. The models are applied to a data set from a Canadian life annuity portfolio. Frees et al. (1996) observed 14,889 policies where one person was male and the other person was female over the period December 29, 1988, through December 31, 1993. The contracts were joint and last-survivor annuities. Frees et al. (1996) observed the date of birth, date of death (if it was applicable), date of contract initiation, and sex of each person. In their article Frees et al. (1996) use the following notation which we will adopt:

· $X$ and $Y$ are the ages at death of him and her, respectively,

· $x$ and $y$ are the contract initiation ages,

· $t_0$ is the time of contract initiation,

· $a := \max(12/29/1988 - t_0, 0)$ : time from contract initiation to the beginning of the observation period,

· $x + a$ and $y + a$ : his and her entry age, respectively,

· $b := 1/1/94 - \max(12/29/88, t_0)$: observation period,

· $T_1 = X - x - a$ : his future lifetime,

· $T_2 = Y - y - a$ : her future lifetime,

· $T_1$ and $T_2$ are only observed if they are both greater than 0.

$\cdot\ j = 1$ or 2

$\cdot\ T_j^* = \min(T_j, b)$,

$$\cdot\ d_j = \begin{cases} 1 & \text{, if the person } j \text{ survives the observation period } (T_j > b) \\ 0 & \text{, if the person } j \text{ dies during the observation period} \end{cases}$$

5.2 Specification of the Model of Frees et al. (1996)

**Lemma 8** *$X \sim F \Rightarrow F(X) \sim U(0,1)$, where U(0,1) is the uniform distribution on the interval (0,1)*

**Proof:**

We have to consider three cases. The first case is that x is non-positive. This implies that $Pr(F(X) \leq x) = 0$. The second case is that $0 < x < 1$. Since $F^{-1}$ is increasing, we have

$$\begin{aligned} Pr(F(X) \leq x) &= Pr(F^{-1}(F(X)) \leq F^{-1}(x)) \\ &= Pr(X \leq F^{-1}(x)) = F(F^{-1}(x)) = x. \end{aligned}$$

The third case is that $x \geq 1$. Thus, we have $Pr(F(X) \leq x) = 1$.

□

We now want to discuss copulas as a starting point for constructing families of bivariate distributions. Upton and Cook (2002) describe a copula as a function that relates a joint cumulative distribution function to the distribution functions of the individual variables. Let $F$ be the multivariate distribution function for the random variables $X_1, ..., X_n$ and let the cumulative distribution function of $X_j$ be $F_j$ for all $j$. Let $U_j$ be defined by $U_j = F_j(X_j)$ for each $j = 1, ..., n$. This implies by Lemma 8 that the marginal distribution of $U_j$ has a continuous uniform distribution on (0,1). Assume that for each $u_j$ there is a unique $x_j$ such that $x_j = F^{-1}(u_j)$ and let the joint distribution function of $U_1, ..., U_n$ be $C$. Then

$$C(u_1, ...u_n) = Pr[U_j < u_j \text{ for all } j] = F[F_1^{-1}(u_1), ..., F_n^{-1}(u_n)]$$

for all $u_1, ..., u_n$ in (0,1), because $U_j < u_j$ is equivalent to $X_j < F^{-1}(u_j)$.

The function C is called the copula. Another - equivalent - formula to show the relationship is:

$$C[F_1(x_1), ..., F_n(x_n)] = F(x_1, ..., x_n).$$

We still haven't defined a copula. Let us give a precise definition as given by Nelsen (1999):

**Definition 13** *A copula is a function $C$ from $[0,1] \times [0,1]$ to $[0,1]$ with the following properties:*

1. *For every u, v in [0,1],*

$$C(u, 0) = 0 = C(0, v)$$

*and*

$$C(u, 1) = u \textit{ and } C(1, v) = v;$$

2. *For every $u_1, u_2, v_1, v_2$ in [0,1] such that $u_1 \leq u_2$ and $v_1 \leq v_2$,*

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

The name "copula" was chosen to emphasize the manner in which a copula "couples" a joint distribution function to its univariate margins (Nelsen, 1999). Frees et al. (1996) use a copula to express bivariate distributions. They consider a random vector $(X, Y)$, where $X$ and $Y$ represent the ages at death of him and her, respectively. The distribution function of $(X, Y)$ is denoted by $H$, where $H(x, y) = Pr(X \leq x, Y \leq y)$ and $F_1$ and $F_2$ denote the respective marginal distribution functions, that is $F_1(x) = H(x, \infty)$ and $F_2(y) = H(\infty, y)$. We observe bivariate distribution functions of the form:

$$H(x, y) = C(F_1(x), F_2(x)), \tag{42}$$

where $C$ is a copula.

Copulas are useful because they provide a link between the marginal distributions and the bivariate distribution. From equation (42) it is obvious that $H$ is determined if $C$, $F_1$ and $F_2$ are known. There are many possibilities for the copula function. We will look at Frank's copula as presented in Frees et al. (1996). This family can be expressed as

$$C(u, v) = \frac{1}{\alpha} \ln[1 + (\exp(\alpha u) - 1)(\exp(\alpha v) - 1)/(\exp(\alpha) - 1)]. \tag{43}$$

**Theorem 8**

$$H(x, y) = \frac{1}{\alpha} \ln[1 + (e^{\alpha F_1(x)} - 1)(e^{\alpha F_2(y)} - 1)/(\exp(\alpha) - 1)] \tag{44}$$

*is a joint distribution function with marginal distributions $F_1(x)$ and $F_2(y)$ when $\alpha$ is unequal to 0.*

**Proof:**

(1) $H(x, y) = F_{XY}(x, y)$ is a distribution function of two future lifetime random

variables:

$$
\begin{aligned}
H(0,0) &= \frac{1}{\alpha}\cdot\ln\left[1+\frac{(\exp(\alpha F_1(0))-1)(\exp(\alpha F_2(0))-1)}{\exp(\alpha)-1}\right] \\
&= \frac{1}{\alpha}\ln\left[1+\frac{(\exp(\alpha\cdot 0)-1)(\exp(\alpha\cdot 0)-1)}{\exp(\alpha)-1}\right] \\
&= \frac{1}{\alpha}\cdot\ln(1) \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
H(\infty,\infty) &= \frac{1}{\alpha}\ln\left[1+\frac{(\exp(\alpha F_1(\infty))-1)(\exp(\alpha F_2(\infty))-1)}{\exp(\alpha)-1}\right] \\
&= \frac{1}{\alpha}\ln\left[1+\frac{(\exp(\alpha)-1)(\exp(\alpha)-1)}{\exp(\alpha)-1}\right] \\
&= \frac{1}{\alpha}\ln(\exp(\alpha)) \\
&= 1
\end{aligned}
$$

(2) The joint p.d.f. is non-negative:

The second derivative of $H = F_{XY}$ with respect to $x$ and $y$ is denoted as $f_{XY}(x,y)$ or as $h$, the derivative of $F_1$ with respect to $x$ is denoted as $f_1$ and the derivative of $F_2$ with respect to $y$ is denoted as $f_2$.

The first derivative of $C(F_1(x),F_2(y))$ with respect to $y$ is:

$$
\begin{aligned}
&\frac{d}{dy}H(x,y) = \frac{1}{\alpha}\cdot\frac{1}{1+(e^{\alpha F_1(x)}-1)(e^{\alpha F_2(y)}-1)/(\exp(\alpha)-1)} \\
&\cdot\quad \frac{\exp(\alpha F_1(x))\cdot\alpha\cdot f_2(y)\cdot\exp(\alpha F_2(y))-\alpha f_2(y)\exp(\alpha F_2(y))}{e^{\alpha}-1} \\
&= \quad \frac{1}{\alpha}\frac{\exp(\alpha F_1(x))\cdot\alpha\cdot f_2(y)\cdot\exp(\alpha F_2(y))-\alpha f_2(y)\exp(\alpha F_2(y))}{e^{\alpha}-1+(e^{\alpha F_1(x)}-1)(e^{\alpha F_2(y)}-1)}
\end{aligned}
$$

The second derivative of $C(F_1(x), F_2(y))$ with respect to $x$ and $y$ is

$$
\begin{aligned}
h(x,y) &= \frac{d^2}{dxdy}H(x,y) = \frac{d}{dx}\left[\frac{d}{dy}H(x,y)\right] \\
&= \frac{\left[e^{\alpha} - 1 + (e^{\alpha F_1(x)} - 1)(e^{\alpha F_2(y)} - 1)\right]}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \\
&\cdot \frac{\left[e^{\alpha F_1(x)} \cdot \alpha \cdot f_1(x) \cdot \alpha \cdot e^{\alpha F_2(y)} \cdot f_2(y)\right]}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \\
&- \frac{\left[e^{\alpha F_1(x)} \cdot \alpha \cdot f_2(y) \cdot e^{\alpha F_2(y)} - \alpha f_2(y) e^{\alpha F_2(y)}\right]}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \\
&\cdot \frac{\left[\alpha e^{\alpha F_1(x)} f_1(x) e^{\alpha F_2(y)} - \alpha f_1(x) e^{\alpha F_1(x)}\right]}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \\
&= \frac{(e^{\alpha} - 1) \cdot \alpha^2 e^{\alpha F_1(x)} e^{\alpha F_2(y)} f_1(x) f_2(y)}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \\
&+ \frac{e^{\alpha F_1(x)} e^{\alpha F_2(y)} \alpha^2 e^{\alpha F_1(x)} e^{\alpha F_2(y)} f_1(x) f_2(y)}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \\
&- \frac{e^{\alpha F_1(x)} \alpha^2 e^{\alpha F_1(x)} e^{\alpha F_2(y)} f_1(x) f_2(y)}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \\
&- \frac{e^{\alpha F_2(y)} \alpha^2 e^{\alpha F_1(x)} e^{\alpha F_2(y)} f_1(x) f_2(y)}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \\
&+ \frac{\alpha^2 e^{\alpha F_1(x)} e^{\alpha F_2(y)} f_1(x) f_2(y)}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2}
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{e^{\alpha F_1(x)} e^{\alpha F_2(y)} \alpha^2 e^{\alpha F_1(x)} e^{\alpha F_2(y)} f_1(x) f_2(y)}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \\
&+ \frac{e^{\alpha F_1(x)} \alpha^2 e^{\alpha F_1(x)} e^{\alpha F_2(y)} f_1(x) f_2(y)}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \\
&+ \frac{e^{\alpha F_2(y)} \alpha^2 e^{\alpha F_1(x)} e^{\alpha F_2(y)} f_1(x) f_2(y)}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \\
&- \frac{\alpha^2 e^{\alpha F_1(x)} e^{\alpha F_2(y)} f_1(x) f_2(y)}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \\
&= \frac{(e^{\alpha} - 1) \alpha^2 e^{\alpha F_1(x)} e^{\alpha F_2(y)} f_1(x) f_2(y)}{\alpha \cdot \left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \\
&= \frac{(e^{\alpha} - 1) \alpha e^{\alpha F_1(x)} e^{\alpha F_2(y)} f_1(x) f_2(y)}{\left[e^{\alpha} - 1 + \left(e^{\alpha F_1(x)} - 1\right)\left(e^{\alpha F_2(y)} - 1\right)\right]^2} \qquad (45)
\end{aligned}
$$

The joint p.d.f is non-negative:

Case 1: $\alpha > 0$: numerator $\geq 0$ and denominator $\geq 0 \Rightarrow$ p.d.f. $\geq 0$

Case 2: $\alpha < 0$: numerator $\geq 0$ and denominator $\geq 0 \Rightarrow$ p.d.f. $\geq 0$.

(3) $F_1(x)$ and $F_2(y)$ are marginal distributions:

$$
\begin{aligned}
H(x, \infty) &= \frac{1}{\alpha} \ln[1 + (e^{\alpha F_1(x)} - 1)(e^{\alpha} - 1)/(e^{\alpha} - 1)] \\
&= \frac{1}{\alpha} \ln[e^{\alpha F_1(x)}] \\
&= F_1(x).
\end{aligned}
$$

$$
\begin{aligned}
H(\infty, y) &= \frac{1}{a} \ln[1 + (e^{\alpha} - 1)(e^{\alpha F_2(y)} - 1)/(e^{\alpha} - 1)] \\
&= \frac{1}{a} \ln[e^{\alpha F_2(y)}] \\
&= F_2(y).
\end{aligned}
$$

$\square$

**Theorem 9** *The parameter $\alpha$ captures the dependence of $X$ and $Y$. $X$ and $Y$ are independent in the limit as $\alpha \to 0$.*

**Proof:**

$f_{XY}(x, y)$ can be expressed as $f_1(x)f_2(y)A(\alpha)B(\alpha)C(\alpha)$, where

$$
\begin{aligned}
A(\alpha) &= \exp[\alpha(F_1(x) + F_2(y))], \\
B(\alpha) &= \frac{\alpha}{e^{\alpha} - 1}, \\
C(\alpha) &= \frac{1}{[1 + \frac{(e^{\alpha F_1(x)} - 1)(e^{\alpha F_2(y)} - 1)}{\exp(\alpha) - 1}]^2}
\end{aligned}
$$

We want to show that

$$
\lim_{\alpha \to 0} f_{XY}(x, y) = f_1(x)f_2(y)
$$

First, we only look at $\lim(A(\alpha)B(\alpha)C(\alpha))$ as $\alpha \to 0$. We have

$$
\lim_{\alpha \to 0} A(\alpha) = \exp(0) = 1
$$

Using the de l'Hospital Rule, we obtain

$$
\lim_{\alpha \to 0} B(\alpha) = \lim_{\alpha \to 0} \frac{1}{e^{\alpha}} = \frac{1}{1} = 1
$$

and $\lim_{\alpha\to 0} C(\alpha)$ depends on the term in the denominator. Using the de l'Hospital Rule, we obtain:

$$\begin{aligned} & \lim_{\alpha\to 0} \frac{(e^{\alpha F_1(x)}-1)(e^{\alpha F_2(y)}-1)}{\exp(\alpha)-1} \\ = & \lim_{\alpha\to 0} \frac{F_1(x)e^{\alpha F_1(x)}(e^{\alpha F_2(y)}-1)}{e^{\alpha}} + \lim_{\alpha\to 0} \frac{F_2(y)e^{\alpha F_2(y)}(e^{\alpha F_1(x)}-1)}{e^{\alpha}} \\ = & \ 0/1 = 0 \end{aligned}$$

Therefore $\lim_{\alpha\to 0} C(\alpha) = \frac{1}{(1+0)^2} = 1$, and $\lim_{\alpha\to 0} f_{XY}(x,y) = f_1(x)f_2(y)$.

□

To completely specify the model, each marginal distribution is assumed to be Gompertz. We will use another parametrization for Gompertz model than the one discussed in chapter 3:

$$\mu(x) = \frac{1}{c}\exp\left[\frac{x-m}{c}\right] \tag{46}$$

With $B = \frac{1}{c}e^{-\frac{m}{c}}$ and $b = \frac{1}{c}$, it can be seen that (46) is only a reparameterized version of the expression for Gompertz used in (22). We are using this parametrization because the parameters are informative. For example, $m$ is the mode of the density (see Lemma 9), and $c$ is approximately the standard deviation.

**Lemma 9** *$m$ is the mode of the density function corresponding to (46).*

**Proof:**

The density function for a newborn is (see (3))

$$\begin{aligned} & f_{T(0)}(x) = {}_xp_0 \cdot \mu(x) \\ = & \exp\left[-\int_0^x \frac{1}{c}\exp\left[\frac{s-m}{c}\right]\right] ds \cdot \frac{1}{c}\exp\left[\frac{x-m}{c}\right] \end{aligned}$$

$$= \frac{1}{c}\exp\left[-\exp\left[\frac{x-m}{c}\right]+\exp\left[\frac{-m}{c}\right]+\left[\frac{x-m}{c}\right]\right].$$

We want to set the first derivative equal to zero using Maple:

Figure 1: Mode of the Density

```
> diff((1/c)*exp((x-m)/c+exp(-m/c)-exp((x-m)/c)),x);
```

$$\frac{\left(\frac{1}{c}-\frac{\mathbf{e}^{\left(\frac{x-m}{c}\right)}}{c}\right)\mathbf{e}^{\left(\frac{x-m}{c}+\mathbf{e}^{\left(-\frac{m}{c}\right)}-\mathbf{e}^{\left(\frac{x-m}{c}\right)}\right)}}{c}$$

```
> solve({(1/c-exp((x-m)/c)/c)*exp((x-m)/c+exp(-m/c)-exp((x-m)/c))/
  c=0},{x});
```

$$\{x=m\}$$

```
>
```

Since the second derivative of the density function evaluated at the point $x = m$ is (using Maple)

$$-\exp(-1+\exp(-m/c))/(c^3) < 0,$$

$m$ is the maximum of the density function. The distribution function of the future lifetime under this parametrization of Gompertz model is

$$\begin{aligned}
F(x) &= 1-\exp\left[-\int_0^x \mu(y)dy\right] \\
&= 1-\exp\left[-\frac{1}{c}e^{-\frac{m}{c}}\int_0^x e^{\frac{y}{c}}dy\right] \\
&= 1-\exp\left[-\frac{1}{c}e^{-\frac{m}{c}}\int_0^{x/c} e^{u}cdu\right] \\
&= 1-\exp\left[-e^{-\frac{m}{c}}(e^{\frac{x}{c}}-1)\right] \\
&= 1-\exp\left[e^{-\frac{m}{c}}(1-e^{\frac{x}{c}})\right]. \qquad (47)
\end{aligned}$$

Our model for modelling the dependence of future lifetimes of coupled lives is specified. We assume the marginals being Gompertz and express the bivariate distribution by Frank's copula. The model has five parameters: $m_1, c_1, m_2, c_2$, $\alpha$.

### 5.3 Maximum-Likelihood-Estimation

#### 5.3.1 Maximum-Likelihood Method in General

We now want to use the maximum-likelihood method to compare several marginal distributions. We will need the derivative of $H$ with respect to $x$, which will be denoted by $H_1$, the derivative of $H$ with respect to $y$ will be denoted by $H_2$ and the second derivative of $H$ with respect to $x$ and $y$ will be denoted by $h$. We develop the likelihood function at first in general, only assuming that the derivatives with respect to $x$ and $y$ of $H(x, y)$ exist. Then we develop it for the model assuming the marginals being Gompertz and expressing the bivariate distributions by Frank's copula. Recall that $T_1$ and $T_2$ are only observed if they are both greater than 0. We define the conditional distribution function of $T_1$ and $T_2$:

$$
\begin{aligned}
& H_T(t_1, t_2) = Pr(T_1 \leq t_1, T_1 \leq t_2 | \, T_1 \text{ and } T_2 \text{ are observed}) \\
= \; & \frac{Pr(0 < T_1 \leq t_1, 0 < T_2 \leq t_2)}{Pr(T_1 > 0, T_2 > 0)} \\
= \; & \frac{Pr(0 < X - x - a \leq t_1, 0 < Y - y - a \leq t_2)}{Pr(T_1 > 0, T_2 > 0)} \\
= \; & \frac{Pr(x + a < X \leq t_1 + x + a, y + a < Y \leq t_2 + y + a)}{Pr(X > x + a, Y > y + a)} \\
= \; & \frac{Pr(X \leq t_1 + x + a, y + a < Y \leq t_2 + y + a)}{1 - [Pr(X < x + a) + Pr(Y < y + a) - Pr(X < x + a, Y < y + a)]} \\
- \; & \frac{Pr(X < x + a, y + a < Y \leq t_2 + y + a)}{1 - [Pr(X < x + a) + Pr(Y < y + a) - Pr(X < x + a, Y < y + a)]} \\
= \; & \frac{Pr(X \leq t_1 + x + a, Y \leq t_2 + y + a) - Pr(X \leq t_1 + x + a, Y \leq y + a)}{1 - [H(x + a, \infty) + H(\infty, y + a) - H(x + a, y + a)]}
\end{aligned}
$$

$$
\begin{aligned}
- & \frac{Pr(X < x+a, Y \le t_2+y+a) - Pr(X < x+a, Y \le y+a)}{1-[H(x+a,\infty)+H(\infty,y+a)-H(x+a,y+a)]} \\
= & \frac{H(t_1+x+a,t_2+y+a) - H(t_1+x+a,y+a)}{1-H(x+a,\infty)-H(\infty,y+a)+H(x+a,y+a)} \\
- & \frac{H(x+a,t_2+y+a) - H(x+a,y+a)}{1-H(x+a,\infty)-H(\infty,y+a)+H(x+a,y+a)}. \qquad (48)
\end{aligned}
$$

Frees at al (1996) point out that this data is *censored* in that most policyholders survived through the end of the observation period. Being *censored* means that he or she survives the end of the observation period, and being *uncensored* means that he or she dies during the observation period. We have to consider four cases: the lifetimes can be both uncensored, the first uncensored and the second censored, the first censored and the second uncensored and both censored.

· Case 1: Both lifetimes are uncensored

We may assume that $t_1 < b$ and $t_2 < b$. We are in the case, where both die during the observation period. Using the notation from section 5.1 we have that $d_1 = 0$ and $d_2 = 0$.

$$
\begin{aligned}
& Pr(T_1^* < t_1, T_2^* < t_2 | \, T_1^* \text{ and } T_2^* \text{ are observed}) \\
= \; & Pr(\min(T_1, b) < t_1, \min(T_2, b) < t_2 | \, T_1 > 0, T_2 > 0) \\
= \; & Pr(T_1 < t_1, T_2 < t_2 | \, T_1 > 0, T_2 > 0) \\
= \; & H_T(t_1, t_2)
\end{aligned}
$$

The contribution to the likelihood-function is the second derivative of $H_T(t_1, t_2)$ with respect to $t_1$ and $t_2$. Using (48) we get:

$$
\frac{h(x+a+t_1, y+a+t_2)}{1-H(x+a,\infty)-H(\infty,y+a)+H(x+a,y+a)} \qquad (49)
$$

· Case 2 : first lifetime uncensored, second censored.

We may assume: $t_1 < b$ and $t_2 \geq b$. We are in the case where he dies during the observation period and she survives the observation period. Therefore we have $d_1 = 0$ and $d_2 = 1$.

$$\begin{aligned}
& Pr(T_1^* < t_1, T_2^* = t_2 | \, T_1^* \, \text{and} \, T_2^* \text{ are observed}) \\
= \; & Pr(\min(T_1, b) < t_1, \min(T_2, b) = b | \, T_1 > 0, T_2 > 0) \\
= \; & Pr(T_1 < t_1, T_2 \geq b | \, T_1 > 0, T_2 > 0) \\
= \; & H_T(t_1, \infty) - H_T(t_1, b)
\end{aligned}$$

The contribution to the likelihood-function is the derivative of this probability with respect to $t_1$ . Using (48) we get:

$$\begin{aligned}
& H_T(t_1, \infty) - H_T(t_1, b) \\
= \; & \frac{H(x+a+t_1, \infty) - H(x+a, \infty)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)} \\
- \; & \frac{H(x+a+t_1, y+a) - H(x+a, y+a)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)} \\
- \; & \frac{H(x+a+t_1, y+a+b) - H(x+a, y+a+b)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)} \\
+ \; & \frac{H(x+a+t_1, y+a) - H(x+a, y+a)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)}
\end{aligned}$$

So, the derivative of this probability with respect to $t_1$ is :

$$\begin{aligned}
& \frac{H_1(x+a+t_1, \infty) - H_1(x+a+t_1, y+a)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)} \\
- \; & \frac{H_1(x+a+t_1, y+a+b) - H_1(x+a+t_1, y+a)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)} \\
= \; & \frac{H_1(x+a+t_1, \infty) - H_1(x+a+t_1, y+a+b)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)} \qquad (50)
\end{aligned}$$

· Case 3 : first lifetime censored, second uncensored.

We may assume: $t_1 \geq b$ and $t < b$. We are in the case, where he survives the observation period and she dies during the observation period. Therefore $d_1 = 1$ and $d_2 = 0$.

$$
\begin{aligned}
& Pr(T_1^* = b, T_2^* < t_2 | \, T_1^* \text{ and } T_2^* \text{ are observed}) \\
= \; & Pr(\min(T_1, b) = b, \min(T_2, b) < t_2 | \, T_1 > 0, T_2 > 0) \\
= \; & Pr(T_1 \geq b, T_2 < t_2 | T_1 > 0, T_2 > 0) \\
= \; & H_T(\infty, t_2) - H_T(b, t_2)
\end{aligned}
$$

The contribution to the likelihood-function is the derivative of this probability with respect to $t_2$. Using (48) we get:

$$
\begin{aligned}
& H_T(\infty, t_2) - H_T(b, t_2) \\
= \; & \frac{H(\infty, y+a+t_2) - H(x+a, y+a+t_2)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)} \\
- \; & \frac{H(\infty, y+a) - H(x+a, y+a)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)} \\
- \; & \frac{H(x+a+b, y+a+t_2) - H(x+a, y+a+t_2)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)} \\
+ \; & \frac{H(x+a+b, y+a) - H(x+a, y+a)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)}
\end{aligned}
$$

So, the derivative of this probability with respect to $t_2$ is:

$$
\begin{aligned}
& \frac{H_2(\infty, y+a+t_2) - H_2(x+a, y+a+t_2)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)} \\
- \; & \frac{H_2(x+a+b, y+a+t_2) - H_2(x+a, y+a+t_2)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)} \\
= \; & \frac{H_2(\infty, y+a+t_2) - H_2(x+a+b, y+a+t_2)}{1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)} \qquad (51)
\end{aligned}
$$

· Case 4 : both lifetimes censored.

We may assume: $t_1 \geq b$ and $t_2 \geq b$. We have $d_1 = 1$ (he survives the observation period) and $d_2 = 1$ (she survives the observation period). The contribution to the likelihood-function is:

$$
\begin{aligned}
& Pr(T_1^* = b, T_2^* = b | \, T_1^* \,\text{and}\, T_2^* \,\text{are observed}) \\
= \;& Pr(\min(T_1, b) = b, \min(T_2, b) = b | \, T_1 > 0, T_2 > 0) \\
= \;& Pr(T_1 \geq b, T_2 \geq b | \, T_1 > 0, T_2 > 0) \\
= \;& Pr(X - x - a \geq b, Y - y - a \geq b | \, T_1 > 0, T_2 > 0) \\
= \;& \frac{Pr(X \geq x + a + b, Y \geq y + a + b)}{Pr(T_1 > 0, T_2 > 0)} \\
= \;& \frac{1 - Pr(X < x + a + b) - Pr(Y < y + a + b)}{Pr(X > x + a, Y > y + b)} \\
+ \;& \frac{Pr(X < x + a + b, Y < y + a + b)}{Pr(X > x + a, Y > y + b)} \\
= \;& \frac{1 - H(x + a + b, \infty) - H(\infty, y + a + b) + H(x + a + b, y + a + b)]}{1 - [Pr(X < x + a) + Pr(Y < y + b) - Pr(X < x + a, Y < y + b)]} \\
= \;& \frac{1 - H(x + a + b, \infty) - H(\infty, y + a + b) + H(x + a + b, y + a + b)}{1 - H(x + a, \infty) - H(\infty, y + b) + H(x + a, y + b)}
\end{aligned} \tag{52}
$$

If we combine these four cases, we can get the likelihood-function for a single observation as :

$$
\begin{aligned}
& L(x, y, t_1, t_2, d_1, d_2, a, b) \\
= \;& [(h(x + a + t_1, y + a + t_2))^{(1-d_1)(1-d_2)} \\
& \cdot(H_1(x + a + t_1, \infty) - H_1(x + a + t_1, y + a + b))^{(1-d_1)d_2} \\
& \cdot(H_2(\infty, y + a + t_2) - H_2(x + a + b, y + a + t_2))^{d_1(1-d_2)}
\end{aligned}
$$

$$\cdot(1 - H(x+a+b,\infty) - H(\infty, y+a+b) + H(x+a+b, y+a+b))^{d_1 \cdot d_2}]/$$
$$[1 - H(x+a,\infty) - H(\infty, y+a) + H(x+a, y+a)] \qquad (53)$$

Thus, the logarithm of the likelihood-function for a single observation is:

$$\begin{aligned}
& \ln[L(x, y, t_1, t_2, d_1, d_2, a, b)] \\
= \ & (1-d_1)(1-d_2) \cdot \ln[h(x+a+t_1, y+a+t_2)] \\
+ \ & ((1-d_1)d_2) \cdot \ln[H_1(x+a+t_1, \infty) - H_1(x+a+t_1, y+a+b)] \\
+ \ & (d_1(1-d_2)) \cdot \ln[H_2(\infty, y+a+t_2) - H_2(x+a+b, y+a+t_2)] \\
+ \ & (d_1 \cdot d_2) \cdot \ln[1 - H(x+a+b, \infty) \\
- \ & H(\infty, y+a+b) + H(x+a+b, y+a+b)] \\
- \ & \ln[1 - H(x+a, \infty) - H(\infty, y+a) + H(x+a, y+a)] \qquad (54)
\end{aligned}$$

The log-likelihood-function for the whole data set is :

$$\sum_{i=1}^{14889} \ln[L(x_i, y_i, t_{1i}, t_{2i}, d_{1i}, d_{2i}, a_i, b_i)]$$

5.3.2 Maximum-Likelihood Method for our Model

To evaluate the maximum-likelihood-estimates for the model assuming the marginals being Gompertz and expressing the bivariate distributions by Frank's copula, we need the derivatives of equation (44)

$$H(x,y) = \frac{1}{a} \ln[1 + (e^{\alpha F_1(x)} - 1)(e^{\alpha F_2(y)} - 1)/(e^{\alpha} - 1)] :$$

$$H_1(x,y) = \frac{f_1(x) e^{\alpha F_1(x)} (e^{\alpha F_2(y)} - 1)}{e^{\alpha} - 1 + (e^{\alpha F_1(x)} - 1)(e^{\alpha F_2(y)} - 1)}$$

and

$$H_2(x,y) = \frac{f_2(x)e^{\alpha F_2(y)}(e^{\alpha F_1(x)}-1)}{e^{\alpha}-1+(e^{\alpha F_1(x)}-1)(e^{\alpha F_2(y)}-1))}$$

and (see equation (45))

$$h(x,y) \quad = \quad \frac{(e^{\alpha}-1)ae^{\alpha F_1(x)}e^{\alpha F_2(y)}f_1(x)f_2(y)}{[e^{\alpha}-1+(e^{\alpha F_1(x)}-1)(e^{\alpha F_2(y)}-1)]^2},$$

where

$$\begin{aligned}
F_j(x) &= 1-\exp[\exp(-m_j/c_j)(1-\exp(x/c_j))] \text{ (see (47))} \\
f_j(x) &= -\exp[\exp(-m_j/c_j)(1-\exp(x/c_j))]\cdot\exp(-m_j/c_j) \\
&\quad \cdot(-1)\exp(x/c_j)\cdot(1/c_j) \\
&= \exp[\exp(-m_j/c_j)(1-\exp(x/c_j))]\cdot\exp((x-m_j)/c_j)\cdot(1/c_j)
\end{aligned}$$

$F_1(x)$ is the distribution function of the future lifetime of the husband and $F_2(x)$ is the distribution function of the future lifetime of the wife, and $m_1$, $c_1$ are the parameters of the distribution function of the husband and $m_2$, $c_2$ are the parameters of the distribution of his wife, $j$ is either 1 or 2.

Frees et al. (1996) get the following maximum-likelihood estimates for their model assuming the marginals being Gompertz and expressing the bivariate distributions by Frank's copula:

$$\begin{aligned}
m_1 &= 85.82 \\
c_1 &= 9.98 \\
m_2 &= 89.40 \\
c_2 &= 8.12 \\
\alpha &= -3.367
\end{aligned} \tag{55}$$

and they get the following maximum-likelihood estimates assuming independence and still using Gompertz marginals:

$$\begin{aligned}
& m_1 = 86.38 \\
& c_1 = 9.83 \\
& m_2 = 92.17 \\
& c_2 = 8.11 \\
& \alpha \text{ is not applicable}
\end{aligned} \tag{56}$$

To be able to interpret the dependence parameter $\alpha$, we should convert it to Spearman's correlation coefficient $\rho_s$. Recall Spearman's correlation coefficient from equation (39)

$$\rho_s = 6Pr[(X_1 - X_2)(Y_1 - Y_3) > 0] - 3. \tag{57}$$

If $C$ is the Frank's copula and the marginals are Gompertz, then Frees et al. (1996) state that:

$$\rho(\alpha) = 1 - 12(D_2(-\alpha) - D_1(-\alpha))/\alpha, \tag{58}$$

where

$$D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{\exp(t) - 1} dt \tag{59}$$

Using Maple we get Figure 2:

Figure 2: Spearman

```
> int(t/(exp(t)-1),t=0..3.367);
                                        1.491950961
> int(t^2/(exp(t)-1),t=0..3.367);
                                        1.702578300
>
```

Thus, we have

$$D_1(-\alpha) = D_1(3.367) = \frac{1}{3.367} \cdot 1.491950961 = 0.443109879$$

and

$$D_2(-\alpha) = D_2(3.367) = \frac{2}{3.367^2} \cdot 1.702578300 = 0.300366059.$$

Hence

$$\rho(\alpha) = \rho(-3.367) = 1 - 12(D_2(3.367) - D_1(3.367))/(-3.367) = 0.49126.$$

If we have independence the correlation is zero (see Theorem 7). Since $\rho_s = 0.49$ this shows a strong statistical dependence (For dependence of the future lifetimes see also section 5.6).

5.4 How Dependence of Joint Lives Effects Annuity Values

We describe as in Frees at al. (1996) the effects of our model of mortality on annuity values. The last-survivor annuity for insureds $x$ and $y$ is an annuity in respect to $(\overline{xy})$. Applying (13), we get that the actuarial present value of this annuity is

$$\ddot{a}_{\overline{xy}} = \sum_{k=0}^{\infty} v^k {}_kp_{\overline{xy}}, \tag{60}$$

where $v$ is the discount factor and $i$ is a constant effective interest rate and

$${}_kp_{\overline{xy}} = 1 - H_T(k,k). \tag{61}$$

$H_T$ is as defined in equation(48).

The maximum likelihood estimates with and without independence are listed in (55) and (56). The annuity values are compared by dividing the annuity values estimated without an independence assumption by those estimated with the independence assumption. Frees et al. (1996) observe this ratio for example for joint and $r$ annuities. These annuities pay a special amount, for example 1 dollar, while both annuitants are alive and $r$ dollars while only one annuitant is alive. Usually $r$ is two-thirds or one-half. In the United States, for these annuities there may be a larger market than for annuities with $r = 1$ because the Employee Retirement Income Security Act (ERISA) mandates all qualified pension plans to offer to qualified beneficiaries a joint and survivor annuity with $r$ at least 50 percent (see Chapter I for details). The actuarial present value for a joint and $r$ annuity is

$$\ddot{a}_{xy}(r) = \sum_{k=0}^{\infty} v^k (r\,{}_kp_x + r\,{}_kp_y - (2r-1)\,{}_kp_{xy}), \tag{62}$$

where ${}_kp_x = 1 - H_T(k,\infty)$ is the conditional probability that a person aged $x$ survives $k$ years, ${}_kp_y = 1 - H_T(\infty,k)$ is the conditional probability that a person aged $y$ survives $k$ years, and ${}_kp_{xy} = {}_kp_x + {}_kp_y - {}_kp_{\overline{xy}} = 1 - H_T(k,\infty) - H_T(\infty,k) + H_T(k,k)$ is the conditional probability that both persons aged $x$ and $y$ survive $k$ years. Explanation of the formula: What does $r{}_kp_x + r{}_kp_y - 2r{}_kp_{xy} + {}_kp_{xy}$ mean? Let's first look at $r{}_kp_x + r{}_kp_y$: He gets $r$ dollars if he survives and she is dead. Similarly, she gets $r$ dollars if she survives and he is dead and they both get $r$ dollars while they are both alive. This implies that we have to subtract $2r$ times the probability that they are both alive and then we still have to add ${}_kp_{xy}$, because they get 1 dollar while they are both alive.

Frees et al. (1996) evaluated the following ratios of dependent to independent joint and $r$ annuity values assuming 5 percent interest and equal annuity age. A ratio of less than one indicates that the annuity values assuming independence are larger than those assuming dependence:

Table 1: Ratios of Dependent to Independent Joint and r Annuity Values

| | r=0 | r=1/4 | r=1/3 | r=1/2 | r=2/3 | r=1 |
|---|---|---|---|---|---|---|
| 50 years | 1.0 | 0.99 | 0.99 | 0.98 | 0.98 | 0.97 |
| 55 years | 1.0 | 0.98 | 0.98 | 0.98 | 0.97 | 0.96 |
| 60 years | 0.99 | 0.98 | 0.98 | 0.97 | 0.96 | 0.95 |
| 65 years | 0.98 | 0.97 | 0.97 | 0.96 | 0.96 | 0.95 |
| 70 years | 0.97 | 0.96 | 0.96 | 0.95 | 0.95 | 0.94 |
| 75 years | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |
| 80 years | 0.89 | 0.91 | 0.92 | 0.93 | 0.94 | 0.95 |

We can see that almost every ratio is less than one. For joint and $r = 1$ annuities, we can see that the ratio is approximately 0.95. This implies that the annuity values for this special annuity are reduced by approximately 5 percent when dependent mortality models are used compared to the standard model of independence.

5.5 Review of Alternative Notation

In this subsection we want to present some alternative notation developed by Carriere (2000). Recall from section 5.1 that 14,889 policies where one person was female and the other was male were observed. Carriere (2000) writes the observations for each policy $k = 1 \ldots$ 14,889 in an observation vector $X_k = (x_{mk}, x_{fk}, t_{mk}, t_{fk}, l_k)'$, where

$\cdot$ $x_{mk}$ is the age of the male at the time of entering the study,

$\cdot$ $x_{fk}$ is the age of the female at the time of entering the study,

$\cdot$ $t_{mk}$ is zero if he survives the observation period and otherwise $t_{mk} + x_{mk}$ is his age at death.

$\cdot$ Similarly, $t_{fk}$ is zero if she survives the observation period and otherwise $t_{fk} + x_{fk}$ is her age at death.

$\cdot$ $l_k$ is the length of time from the date of entering the study to the end of the study. Usually $l_k$ =12/31/93 - 29/12/88 = 5 years +2 days = 5 years +2/365 years =5.055 years.

$\cdot$ Let $\theta$ denote the vector of parameters.

$\cdot$ $S_{mf}(t_m, t_f | \theta) = Pr(T(x_m) > t_m, T(x_f) > t_f)$ denotes the bivariate survival function of the time-at-death random variables $T(x_m)$ and $T(x_f)$.

$\cdot$ $T(x_m)$ is the future lifetime of a person currently aged $x_m$.

$\cdot$ $S_1^{mf}$ denotes the partial derivative of $S_{mf}$ with respect to $t_m$,

$\cdot$ $S_2^{mf}$ denotes the partial derivative of $S_{mf}$ with respect to $t_f$, and

$\cdot$ $S_{12}^{mf}$ denotes the partial derivative of $S_{mf}$ with respect to $t_m$ and $t_f$.

$\cdot$ The likelihood function is denoted as $L(\theta)$. Its representation is given in terms of survival functions and not in terms of cumulative distributions functions as in Frees et al. (1996).

$$L(\theta) = \prod_{k=1}^{14{,}889} [L_k^{dd}]^{\delta_k^m \delta_k^f} [L_k^{ds}]^{\delta_k^m (1-\delta_k^f)} [L_k^{sd}]^{(1-\delta_k^m)\delta_k^f} [L_k^{ss}]^{(1-\delta_k^m)(1-\delta_k^f)}, \tag{63}$$

where

$$\begin{aligned}
\delta_k^m &= \begin{cases} 1 & \text{, if he dies} \\ 0 & \text{, if he survives} \end{cases} \\
\delta_k^f &= \begin{cases} 1 & \text{, if she dies} \\ 0 & \text{, if she survives} \end{cases} \\
L_k^{dd} &= S_{12}^{mf}(t_{mk}, t_{fk}|\, \theta), \\
L_k^{ds} &= -S_1^{mf}(t_{mk}, l_k|\, \theta), \\
L_k^{sd} &= -S_2^{mf}(l_k, t_{fk}|\, \theta), \\
L_k^{ss} &= S^{mf}(l_k, l_k|\, \theta),
\end{aligned}$$

The marginal functions of $S_{mf}(t_m, t_f|\, \theta)$ will be denoted as $S_m, S_f$ and they are defined as $S_m(t|\, \theta) = S_{mf}(t, 0|\, \theta)$ and $S_f(t|\, \theta) = S_{mf}(0, t|\, \theta)$. Since

$$S_m(t|\, \theta) = Pr(T(x_m) > t) = {}_tp_{x_m} = \frac{{}_{t+x_m}p_0}{{}_{x_m}p_0},$$

only ${}_xp_0$ needs to be defined for the marginals.

5.6 Dependence of his and her Future Lifetime

There are different possibilities to summarize the dependence between two random variables by a single measure. If we suppose that the death of a man, aged $x_m$, has no effect on the death or survival of his wife, aged $x_f$ (or vice versa), then the two lives are not associated. This means, stochastically, that $T(x_m)$ and $T(x_f)$ are independent, i.e. $Pr(T(x_m) \leq t_m, T(x_f) \leq t_f) = Pr(T(x_m) \leq t_m)Pr(T(x_f) \leq t_f)$ for all $t_m, t_f$. In this case a usual measure of association will be zero. Carriere and Chan (1986) defines a pair of lives *in agreement* if a long life for one is associated with a long life for the other. In this case, the measure will be positive. He defines the two lives to be *in perfect agreement* if there exists a strictly increasing function $I(\cdot)$ such that $T(x_m) = I(T(x_f))$. Similarly, he defines the lives to be *in disagreement* if a short life for one is associated with a long life for the other. In this case, the measure will be negative. And he defines the lives to be *in perfect disagreement* if there exists a strictly decreasing function $D(\cdot)$ such that $T(x_f) = D(T(x_m))$. Usually, the measure of association is the linear correlation coefficient. Let $\mu_{x_m} = E[T(x_m)]$, $\sigma^2_{x_m} = Var[T(x_m)]$ and similarly $\mu_{x_f} = E[T(x_f)]$, $\sigma^2_{x_f} = Var[T(x_f)]$.

**Definition 14** *The linear correlation coefficient is:*

$$\begin{aligned}
Cor[T(x_m), T(x_f)] &= \frac{Cov[T(x_m), T(x_f)]}{\sigma_{x_m}\sigma_{x_f}} \\
&= \frac{E[T(x_m)T(x_f)] - \mu_{x_m}\mu_{x_f} - \mu_{x_m}\mu_{x_f} + \mu_{x_m}\mu_{x_f}}{\sigma_{x_m}\sigma_{x_f}} \\
&= E\left[\frac{T(x_m)T(x_f) - \mu_{x_m}T(x_f) - \mu_{x_f}T(x_m) + \mu_{x_m}\mu_{x_f}}{\sigma_{x_m}\sigma_{x_f}}\right] \\
&= E\left[\frac{(T(x_m) - \mu_{x_m})(T(x_f) - \mu_{x_f})}{\sigma_{x_m}\sigma_{x_f}}\right]
\end{aligned}$$

Other measures are Spearman's correlation coefficient and Kendall's population correlation coefficient, which are defined as (see Definition 10 and Definition 12)

$$\rho_s = 6Pr[(X_{m1} - X_{m2})(X_{f1} - X_{f3}) > 0] - 3 \tag{64}$$

and

$$\tau = 2Pr[(X_{m2} - X_{m1})(X_{f2} - X_{f1}) > 0] - 1. \tag{65}$$

If $T_{x_m}$ and $T_{x_f}$ are independent, both measures are zero (see Theorem 6 and Theorem 7).

Assume that the marginal distributions are Gompertz with paramters $m_m = 86.37, \sigma_m = 9.83, m_f = 92.17, \sigma_f = 8.11$ (see (55)). Let $G_{i,k}$ denote the known continuous distribution function of $T(x_{i,k})$, given that the death occur during the observation period ($i = m, f$ and $k = 1, ..., 229$; only the information in the 229 policies is used where both persons died during the observation period because the censored observations should be neglected.).

$$\begin{aligned} G_{i,k}(t_{i,k}) &= Pr(T(x_{i,k}) \leq t_{i,k} | T(x_{i,k}) \leq l_k) \\ &= \frac{Pr(T(x_{i,k}) \leq t_{i,k})}{Pr(T(x_{i,k}) \leq l_k)} \\ &= \frac{1 - S^i(t_{i,k}|\theta)}{1 - S^i(l_k|\theta)}. \end{aligned}$$

$G_{i,k}(T(x_{i,k})) =: U_{i,k} \sim U(0,1)$ (see Lemma 8). Under the null hypothesis of independence of $T(x_m)$ and $T(x_f)$, the pairs $(U_{mk}, U_{fk})$ for $k = 1, ..., 229$ are independent and identically distributed with a common copula $C(u, v) = uv$. This means that $(U_{11}, U_{21}), ...(U_{1,229}, U_{2,229})$ is a random sample from a continuous bivariate population (see section (4.2) for more details about these tests)

Let $R_{i,k}$ denote the rank of $U_{i,k}$. Then the estimate of Spearman's correlation coefficient is

$$\hat{\rho}_s = \frac{\sum_{k=1}^{229}[R_{mk} - \frac{229+1}{2}][R_{fk} - \frac{229+1}{2}]}{229(229^2-1)/12} = 0.415.$$

(For a justification of $\hat{\rho}_s$ see (41)), the estimate of Kendall's Tau $\hat{\tau}$ is 0.325.

Under the assumption of independence and for a large sample size $\hat{\rho}_s$ is asymptotically normal with mean of zero and variance of $\frac{1}{229-1}$ (Lemma 7). Since $\hat{\tau} = \frac{2K}{n(n-1)} = 0.325$, this implies that $K = 8484.45$. Under independence and for a large sample size, $K$ is asymptotically normal with a mean of zero and a variance of $\frac{n(n-1)(2n+5)}{18}$ (see Lemma 4). We have $|\frac{\hat{\rho}_s}{\sqrt{1/(n-1)}}| = \sqrt{228}\cdot\hat{\rho}_s = 6.266 > z_{0.025} = 1.96$, where $z_{0.025}$ is the 0.025-quantile of the standard normal distribution, and $|K^*| = \frac{K}{\sqrt{n(n-1)(2n+5)/18}} = 7.32 > z_{0.025} = 1.96$. Thus the null hypothesis of independence of $T(x_m)$ and $T(x_f)$ can be rejected. The result from this method must be used with caution because we are assuming that the marginals are known (Gompertz) and we only use 229 observations.

5.7 Alternative Bivariate Models

5.7.1 Alternative Marginals

In this section the Gompertz law for the marginals will be compared to the Inverse-Gompertz, Lognormal, Weibull and Gamma models. The survival functions for the five models:

$$\text{Gompertz: } {}_xp_0 = \exp\left[e^{-\frac{m}{\sigma}}(1 - e^{\frac{x}{\sigma}})\right] \tag{66}$$

$$\text{Weibull: } {}_xp_0 = \exp(-(x/m)^{m/\sigma}) \tag{67}$$

$$\text{Gamma: } {}_xp_0 = \int_x^\infty \frac{(zm/\sigma^2)^{m^2/\sigma^2}}{z\Gamma(m^2/\sigma^2)} \exp(-zm/\sigma^2)dz \tag{68}$$

$$\text{Lognormal: } {}_xp_0 = \int_{m/\sigma \ln(x/m)}^\infty \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)dz \tag{69}$$

$$\text{Inverse Gompertz: } {}_xp_0 = \frac{1-\exp(-\exp(-\frac{x-m}{\sigma}))}{1-\exp(-\exp(\frac{m}{\sigma}))} \tag{70}$$

The paramatrization may be unfamiliar, but the parameters are informative. We already used the parametrization for the Gompertz model in equation (46). In all cases $m > 0$ is a location parameter and $\sigma > 0$ is a dispersion parameter (Carriere, 2000).

5.7.1.1 Maximum-Likelihood

Now we want to find the maximum-likelihood estimates for $(m_1, \sigma_1)$ and $(m_2, \sigma_2)$ denoted by $(\hat{m}_1, \hat{\sigma}_1)$ and $(\hat{m}_2, \hat{\sigma_2})$. Estimating the male and female marginals separately is equivalent to estimating the marginals under an independent bivariate model. This means that $S^{mf}(t_m, t_f \mid \theta) = S^m(t_m \mid \theta)S^f(t_f \mid \theta)$. Let $\hat{\theta}$ denote the maximum likelihood estimator of $\theta$. Carriere's results for using these five models as marginals for an independent bivariate model are listed in the following table:

Table 2: Maximum-Likelihood Estimates

| | $\hat{m}_m$ | $\hat{\sigma}_m$ | $\hat{m}_f$ | $\hat{\sigma}_f$ | $-\ln[L(\hat{\theta})]$ |
|---|---|---|---|---|---|
| Gompertz | 86.37 | 9.83 | 92.16 | 8.11 | 10033.8 |
| Weibull | 86.72 | 10.11 | 92.99 | 9.26 | 10047.2 |
| Gamma | 84.47 | 13.23 | 93.36 | 15.39 | 10113.2 |
| Lognormal | 83.94 | 13.64 | 93.28 | 16.64 | 10130.9 |
| Inverse-Gompertz | 79.51 | 14.31 | 89.40 | 18.53 | 10180.5 |

We can observe that $\hat{m}_m < \hat{m}_f$ for all models; implying that the average age of a male is less than the average age of a female for every model. The smallest value for $-\ln[L(\hat{\theta})]$ is for the Gompertz model. This implies that the Gompertz model is the best model of the five considered.

5.7.1.2 Generalized Likelihood Ratio Tests

Since the Log-Likelihood-values of the Gompertz and the Weibull model are very close, we want to test the null hypothesis $H_0$: *Weibull* versus the alternative hypothesis $H_1$: *Gompertz.* Since the hypotheses are non-nested, we first have to look at another test: $H_0$: *Weibull* versus $H_1$: *mixture of Weibull and Gompertz.* Let $Q := pW + (1-p)G$, $0 \leq p \leq 1$, denote a mixture of a Weibull and a Gompertz density. The paramters of $Q$ are two paramters of $G$ and two paramters of $W$ and the paramter $p$, so that the number of paramters of $Q$ is 5. Testing the hypothesis $H_0$: *Weibull* versus the alternative hypothesis $H_1$: $Q$ is equivalent to testing the hypothesis $H_0$: $p = 1$ versus the alternative hypothesis $H_1$: $p \neq 1$. Let $L_Q(\hat{\theta}_Q)$ denote the likelihoodfunction for $Q$ evaluated at the maximum likelihood estimate $\hat{\theta}_Q$ and let $L_W(\hat{\theta}_W)$ denote the likelihoodfunction for $W$ evaluated at the maximum likelihood estimate $\hat{\theta}_Q$. The test statistic is

$$Y := -2\ln[\frac{L_W(\hat{\theta}_W)}{L_Q(\hat{\theta}_Q)}] \sim \chi^2(5-2) = \chi^2(3)$$

(see section 4.1). The test, which rejects $H_0$ in favor of $H_1$: Q whenever $Y > \chi^2(3, \alpha)$, has a significance level of approximately $\alpha$ for a large sample size. However, we don't want to test the hypothesis $H_0$: *Weibull* versus the alternative hypothesis $H_1$:$Q$, instead we want to test the hypothesis $H_0$:*Weibull* versus the alternative hypothesis $H_1$: *Gompertz.* Let $L_G(\hat{\theta}_G)$ denote the likelihoodfunction of the Gompertz model evaluated at the maximum likelihood estimate $\hat{\theta}_G$ . Let

$$T = -2\ln\left[\frac{L_W(\hat{\theta}_W)}{L_G(\hat{\theta}_G)}\right].$$

**Lemma 10**

$$Pr(T > \chi^2(3, a)) < Pr(Y > \chi^2(3, a)) \approx \alpha.$$

**Proof:**

Note that

$$L_Q(\hat{\theta}_Q) = \sup\{f(x_1, ..., x_n; p, m_W, \sigma_W, m_G, \sigma_G) : (p, m_W, \sigma_W, m_G, \sigma_G) \in \theta\},$$

where $m_W$ and $\sigma_W$ are the parameters of the Weibull distribution and $m_G$ and $\sigma_G$ are the parameters of the Gompertz distribution. Besides note that

$$L_G(\hat{\theta}_G) = \sup\{f(x_1, ..., x_n; p, m_W, \sigma_W, m_G, \sigma_G) : (p = 1, 0, 0, m_G, \sigma_G) \in \theta\}.$$

This implies that we take the supremum over a larger set for $L_Q(\hat{\theta}_Q)$ than for $L_G(\hat{\theta}_G)$. Thus, we have that $L_Q(\hat{\theta}_Q)$ is greater than or equal to $L_G(\hat{\theta}_G)$. That is equivalent to

$$\begin{aligned} & \ln[L_Q(\hat{\theta}_Q)] \geq \ln[L_G(\hat{\theta}_G)] \\ \Leftrightarrow\ & -2\ln[L_W(\hat{\theta}_W)] + 2\ln[L_Q(\hat{\theta}_Q)] \geq -2\ln[L_W(\hat{\theta}_W)] + 2\ln[L_G(\hat{\theta}_G)] \\ \Leftrightarrow\ & Y \geq T. \end{aligned}$$

Hence we have

$$Pr(T > \chi^2(3, a)) < Pr(Y > \chi^2(3, a)) \approx \alpha$$

for a large sample size.

□

This implies that if we reject $H_0$: Weibull in favor of $H_1$: Gompertz whenever $T > \chi^2(3, \alpha)$ this test has a level of significance of approximately $\alpha$.

Let's look at the data set (Carriere, 2000):

Table 3: Log-Likelihood Function Evaluated at the Log-Likelihood Estimates

| | $-lnL(\hat{\theta})$ |
|---|---|
| Gompertz | 10033.8 |
| Weibull | 10047.2 |

Thus, the test statistic for the asymptotic generalized ratio test (see section 4.1.2)

$$\begin{aligned} T &= -2\ln[\frac{L_W(\hat{\theta}_W)}{L_G(\hat{\theta}_G)}] = -2\ln[L_W(\hat{\theta}_W)] - (-2\ln[L_G(\hat{\theta}_G)]) \\ &= 2 \cdot 10,047.2 - 2 \cdot 10,033.8 = 26.8 > \chi^2(3, 0.05) = 7.81. \end{aligned}$$

This implies that $H_0$ should be rejected.

5.7.2 Alternative Families

In section 5.2 a model is specified using Gompertz marginals and Frank's copula. In the previous section we have seen that Gompertz's law for the marginals seems to be a good fit. In this section the family due to the Linear-Mixing-Frailty as bivariate distribution is compared to the families due to Frailty, Normal, Generalized Frank, Generalized Normal, Frank and Correlated Frailty. The copulas of these models are listed below:

Frank (F):

$$C(u,v) = \frac{1}{\alpha}\ln\left[\frac{\exp(-\alpha)-1}{\exp(-\alpha(u+v)) - \exp(-\alpha \cdot u) - \exp(-\alpha \cdot v) + \exp(-\alpha)}\right],$$
$$\alpha \in \Re, \alpha \neq 0$$

Frailty (Fr):

$$C(u,v) = \max(0, (u^{1-e^{\alpha}} + v^{1-e^{\alpha}} - 1)^{\frac{1}{1-e^{\alpha}}}),\ \alpha \in \Re, \alpha \neq 0$$

Normal (N):

$$C(u,v) = H(\Phi^{-1}(u), \Phi^{-1}(v)|\, \rho),\ -1 < \rho < 1,$$

where $\Phi^{-1}$ is the inverse function of $\Phi$, that is $\Phi^{-1}(\Phi(t)) = t$ and

$$\Phi(t) = \int_{-\infty}^{t} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

$$H(x,y|\, \rho) = \int_{-\infty}^{x} \int_{-\infty}^{y} \frac{\exp\Big[ -0.5(z_1^2 - 2\rho z_1 z_2 + z_2^2)/(1-\rho^2) \Big]}{2\pi\sqrt{1-\rho^2}} dz_2 dz_1$$

The three models already presented are one-parameter families of copulas. We want to describe how a multiple-parameter family of copulas can be constructed as in Carriere (2000).

**Lemma 11** *Let $g(\cdot)$ be an increasing function with $g(0) = 0$ and $g(1) = 1$. Assume that $g(\cdot)$ has an inverse $g^{-1}(\cdot)$ such that $g^{-1}(g(u)) = u$. If $C(u,v)$ is a copula then $g^{-1}[C(g(u), g(v))]$ is also a copula.*

If $g(u) = u^{\xi}$, $\xi > 0$, then $g(\cdot)$ is an increasing function with $g(0) = 0^{\xi} = 0$, since $\xi > 0$ and thus $\xi \neq 0$, and $g(1) = 1^{\xi} = 1$. The inverse of $g(\cdot)$ is $g^{-1}(u) = u^{1/\xi}$. Hence the preliminaries of Lemma 11 are fulfilled. Using the technique and Frank's copula we obtain a Generalized Frank model, defined as follows:

Generalized Frank:

$$\begin{aligned} C(u,v) &= [C(u^{\xi}, v^{\xi})]^{1/\xi} \\ = & \left[\frac{1}{\alpha}\ln\left[\frac{\exp(-\alpha)-1}{\exp(-\alpha(u^{\xi}+v^{\xi}))-\exp(-\alpha\cdot u^{\xi})-\exp(-\alpha\cdot v^{\xi})+\exp(-\alpha)}\right]\right]^{1/\xi}, \\ & \alpha \in \Re, \alpha \neq 0 \end{aligned}$$

Correlated Frailty (CF):

$$C(u,v) = \frac{(uv)^{1-p}}{(u^{-\alpha}+v^{-\alpha}-1)^{p/\alpha}}$$

Linear-Mixing Frailty (LMF):

$$C(u,v) = (1-p)uv + p(u^{-\alpha}+v^{-\alpha}-1)^{-1/\alpha},\ 0 \leq p \leq 1,\ \alpha > 0$$

5.7.2.1 Maximum-Likelihood

We will use as in Carriere (2000) the Maximum-Likelihood method to compare these copulas and the results for the maximum likelihood estimates and the likelihood function evaluated at the maximum likelihood estimates are shown in the following table (following Carriere, 2000):

Table 4: Maximum-Likelihood Estimates for Various Bivariate Models

| | $\hat{m}_m$ | $\hat{\sigma}_m$ | $\hat{m}_f$ | $\hat{\sigma}_f$ | - | - | $\hat{\rho}_s$ | $-lnL(\hat{\theta})$ |
|---|---|---|---|---|---|---|---|---|
| LMF | 86.5 | 10.1 | 92.5 | 8.39 | $\hat{\alpha} = 2.41$ | $\hat{\rho} = 0.339$ | 0.325 | 9947.2 |
| CF | 86.5 | 10.1 | 92.3 | 8.29 | $\hat{\alpha} = 2.41$ | $\hat{\rho} = 0.348$ | 0.275 | 9947.2 |
| GF | 86.5 | 10.1 | 92.2 | 8.16 | $\hat{\alpha} = 1.54$ | $\xi = 3.06$ | 0.249 | 9949.1 |
| F | 86.5 | 10.2 | 92.5 | 8.31 | $\hat{\alpha} = 3.04$ | - | 0.454 | 9951.1 |
| N | 86.5 | 10.1 | 92.3 | 8.20 | $\hat{\rho} = 0.326$ | - | 0.313 | 9953.9 |
| Fr | 86.6 | 10.5 | 92.8 | 8.51 | $\hat{\alpha} = 1.00$ | - | 0.639 | 9958.3 |
| Ind. | 86.4 | 9.83 | 92.2 | 8.11 | - | - | 0 | 10033.8 |

Since the Linear-Mixing Frailty and the Correlated Frailty have the largest value for the statistic $\ln[L(\hat{\theta})]$, they seem to be good models.

5.7.2.2 Generalized Likelihood Ratio Tests

(See section 4.2 for more details about likelihood ratio tests) We want to test

· a) $H_0$: Generalized Frank, versus the alternate $H_1$: Linear-Mixing Frailty

· b) $H_0$: Correlated Frailty, versus the alternate $H_1$: Linear-Mixing Frailty

· c) $H_0$: Frank, versus the alternate $H_1$: Linear-Mixing Frailty

· d) $H_0$: Normal, versus the alternate $H_1$: Linear-Mixing Frailty

· e) $H_0$: Frailty, versus the alternate $H_1$: Linear-Mixing Frailty

· f) $H_0$: Independent Model, versus the alternate $H_1$: Linear-Mixing Frailty

· a) The test statistic for this test is

$$\begin{aligned} T &= -2\ln[\frac{L_{GF}(\hat{\theta}_{GF})}{L_{LMF}(\hat{\theta}_{LMF})}] \\ &= -2[\ln(L_{GF}(\hat{\theta}_{GF})) - \ln(L_{LMF}(\hat{\theta}_{LMF}))], \end{aligned}$$

where $L_{GF}(\hat{\theta}_{GF})$ denotes the likelihood function for the generalized Frank evaluated at the Maximum-Likelihood estimate $\hat{\theta}_{GF}$ and $L_{LMF}(\hat{\theta}_{LMF})$ is the likelihood function for the Linear Mixing Frailty evaluated at the Maximum-Likelihood estimate $\hat{\theta}_{LMF}$. Since the hypotheses are non-nested, we use the testing methodology as in section 5.7.1.2. The number of degrees of freedom is 3 (Carriere, 2000). Plugging in the values from the previous table we have

$$T = -2\ln[-9949.1 + 9947.2] = 3.8 < \chi^2(3, 0.05) = 7.81.$$

This implies that $H_0$: Generalized Frank should not be rejected.

b) through f) are very similar. We are only looking at the results and using the degrees of freedom in Carriere, 2000.

· b) $T = -2\ln\left[\frac{L_{CF}(\hat{\theta}_{CF})}{L_{LMF}(\hat{\theta}_{LMF})}\right] = -2\ln[-9947.2 + 9947.2] = 0 < \chi^2(1, 0.05) = 3.84.$

This implies that $H_0$: Correlated Frailty is not rejected.

· c) $T = 7.8 > \chi^2(2, 0.05) = 5.99.$

This implies that $H_0$: Frank is rejected.

· d) $T = 13.4 > \chi^2(2, 0.05) = 5.99.$

This implies that $H_0$: Normal is rejected.

· e) $T = 22.2 > \chi^2(1, 0.05) = 3.84.$

This implies that $H_0$: Frailty is rejected.

· f) $T = 173.2 > \chi^2(2, 0.05) = 5.99.$

This implies that $H_0$: Independent model is rejected.

These tests imply that the Linear Mixing Frailty is significantly better than all the other models except the Correlated Frailty and the Generalized Frank model. The independent and Frailty models are nested within the Linear Mixing Frailty model, the generalized Frank, Frank and Normal are non-nested. Since the statistic $\ln[L(\hat{\theta})]$ of the Linear Mixing Frailty is larger than or equal to the statistic of the other models, we can conclude that the Linear Mixing Frailty is the preferred model. Besides it is less complicated than the generalized Frank and the correlated Frailty models.

## REFERENCES


Allen Jr., Everett T., Melone, Joseph T. and Rosenbloom, Jerry S. 2003. *Pension Planning* 9th ed. McGraw-Hill Irwin.

Bowers Jr., Newton L., Gerber, Hans U., Hickman, James C., Jones, Donald A. and Nesbitt, Cecil J. 1997. *Actuarial Mathematics* 2nd ed. Schaumburg, IL: The Society of Actuaries.

Boyers, Judith T. 1986. *Pension in Perspective* 2nd ed. Cincinnati, Ohio: The National Underwriter Company.

Carriere, Jacques F. and Chan, Lai K. 1986. "The bounds of bivariate distributions that limit the value of last-survivor annuities". *Transactions: Society of Actuaries XLIV, p. 77-99.*

Carriere, Jacques F. 2000. "Bivariate survival models for coupled lives". *Scandinavian Actuarial Journal* 2000(1):, 17-32.

Everitt, B. S. 2002. *Dictionary of statistics* 2nd ed. Cambridge, United Kingdom: University Press.

Frees, Edward W., Carriere, Jacques F. and Valdez, Emiliano 1996. "Annuity valuation with dependent mortality". *Journal of Risk and Insurance* 63(2): 229-261.

Fisz, M. 1963. *Probability Theory and Mathematical Statistics* 3rd ed. Malabor, FL: Robert E. Krieger Publishing Company, Inc.

Gajek, Leslaw and Ostaszewski, Krzysztof 2002, *Plany Emerytalne: Zarzadzanie Aktywami i Zobowiazaniami*, Poland, Warsaw: Wydawnictowa Naukowo-Techniczne.

Gajek, Leslaw and Ostaszewski, Krzysztof upcoming in 2003, *Pension Plans and*

*Their Asset-Liability Management*, Massachusetts, Boston: Academic Press.

Hallman, G. Victor and Hamilton, Karen L. 1994. *Personal Insurance: Life, Health, and Retirement* 1st ed. Malvern, Pennsylvania: American Institute for CPCU.

Higgins, Tim 2003. "Mathematical Models of Mortality". *http://acpr.edu.au/ Publications/Mortality%20talk%20-%20mathematical%20equations.pdf*. Canberra, Australia.

Hollander, Myles and Wolfe, Douglas A. 1999. *Nonparametric Statistical Methods* 2nd ed. New York, NY: John Wiley & Sons, Inc.

Kruskal, William H. 1958. "Ordinal measure of association". *Journal of the American Statistical Association* 53: 814-861.

Mood, A. M., Graybill, F. A. and Boes, D. C. 1974. *Introduction to the theory of statistics* 3rd ed. New York, NY: McGraw-Hill.

Nelsen, Roger B. 1999. *An Introduction to Copulas*. New York, NY: Springer.

Rejda, George E. 2003. Principles of Risk Management and Insurance 8th ed. Addison Wesley.

Seton Hall University 2003. *http://www.shu.edu/projects/reals/numser/proofs/ sumparts.html*. South Orange, New Jersey.

Upton, Graham and Cook, Ian 2002. *Dictionary of Statistics*. New York: Oxford University Press Inc.

Zwillinger, Daniel 1992. *Handbook of Integration*. London, England: Jones and Bartlett Publishers, Inc.