# Supplementary Material from New Syllabus

October 27, 2013

The new syllabus included 3 new main learning objectives which were not included as of the Second Edition of our text. In this document, we will cover the new material with plenty of examples and exercise problems. We will also mention which chapter of our book each topic corresponds to. It is recommended to read the indicated chapter first as a prerequisite.

# 1    A.8 (Prerequisite: Chapter 2)

This first topic requires a conceptual understanding of two types of extreme value distributions. It is unlikely that any problems involving these distributions will require computations.

Extreme value distributions are used to model things that are "worst-case", in some sense. For example, our book mentions statistics like value-at-risk (VaR) and conditional tail expectation (CTE) which are examples of extreme values. There are 2 types of extreme value distributions you should know, and we describe them below.

1. **The maximum observation from a random sample.**
   This is equivalent to $VaR_{100\%}(x)$. The Fréchet distribution (more familiarly known as, and included in your formula sheet as, the inverse Weibull distribution) is most often used for modeling this random variable. One reason is that $VaR_{100\%}(x)$ tends to have a heavy tail, which is a characteristic of the inverse Weibull distribution.

2. $\mathbf{E}[(\mathbf{X} - \mathbf{d})_+]$ **as** $\mathbf{d} \to \infty$.
   In an insurance setting, the limiting distribution of this mean excess loss random variable is modeled using the exponential distribution if $X$ is light-tailed, and is modeled via the Pareto distribution if $X$ is heavy-tailed.

   The reason why insurance companies care about this distribution is that some insurance policies have high deductibles (i.e. $d$ is large), and there is a need to understand their average payouts in this setting.

# 2    G (Prerequisite: Chapters 8 and 9)

Chapter 9 already described ways to estimate decrement probabilities for ordinary samples. In the case where the sample size is large, there are other ways to perform calculations. Recall that in general, the conditional one-period decrement probability $q_j$ for the period $[c_j, c_{j+1})$ is given by

$$q_j = \frac{S(c_j) - S(c_{j+1})}{S(c_j)}$$

We now want to be able to estimate $q_j$ given large sample data. To do so, we start with a bit of notation. Let $d_j$ be the number of observed decrements in the interval, and let $e_j$ be the exposure in the interval. Our estimates for $q_j$ will make use of these two variables.

## 2.1 Individual Data

In general, with individual data, there are two methods that can be used to estimate decrement probabilities, each of which differs on the definition of $e_j$, a measure of exposure.

1. **Exact exposure method**
   We set $e_j$ to be the exact time observed within the age interval. When an individual's policy terminates due to a decrement, his or her exposure ends at the exact age of death. We then have:

$$\hat{q}_j = 1 - \exp\left(-\frac{(c_{j+1} - c_j)d_j}{e_j}\right)$$

2. **Actuarial exposure method**
   The exposure period $e_j$ is rounded up to the end of the age interval. We then estimate via:

$$\hat{q}_j = \frac{(c_{j+1} - c_j)d_j}{e_j}$$

In the examples given below, we are working with consecutive integral ages, meaning that $c_{j+1} - c_j = 1$.

**Example 2.1.** Consider the following dataset for 6 individuals:

| ID | Date Joined | Birthdate | Date Left | Reason |
|----|-------------|-----------|-----------|--------|
| 1  | 1/1/03      | 5/1/73    | 11/1/11   | Lapse  |
| 2  | 7/1/02      | 6/1/73    | 12/1/11   | Death  |
| 3  | 5/1/02      | 6/1/73    | 12/1/12   | Death  |
| 4  | 8/1/03      | 9/1/74    | 7/1/12    | Lapse  |

We realize that 4 insureds does not make a large sample size, but in order to illustrate the calculations on a more fine-grain detail, we are using this smaller dataset.

Estimate death probabilities $q_j$ at all integral ages using both the exact exposure method and the actuarial exposure method.

**Answer** We begin by calculating the ages for each event (dates joined and left) for each insured. Individual 1 joined at age 29-8 (a concise notation for 29 years and 8 months of age) and left at age 38-6. Individual 2 joined at age 29-1 and left at age 38-6. Individual 3 joined at age 28-11 and left at age 39-6. Finally, individual 4 joined at age 28-11 and left at age 37-10. There was a death between age 38 and 39 as well as a death between age 39 and 40. So these are the only two decrement probabilities we have to calculate (all others are 0).

1. Exact exposure
   We start by calculating $e_{38}$ and $e_{39}$.

   For $e_{38}$, individuals 1 and 2 contributes 6/12 each, individual 3 contributes 1 (a full year, since he leaves after age 39), and individual 4 contributes 0 (since he leaves before age 38). Thus, $e_{38} = (6 + 6 + 12 + 0)/12 = 2$. Similarly $e_{39} = 6/12 = 0.5$, since only individual 3 was in the policy for any part of ages 39-40. $d_j = 1$ for both time intervals. Thus, $\hat{q}_{38} = 1 - \exp(-1/2) = 0.3935$ and $\hat{q}_{39} = 1 - \exp(-1/0.5) = 0.8647$.

2. Actuarial exposure
   For actuarial exposure, we will need to recalculate the exposure contributions for individuals 2 and 3 (who have died). For calculating exposure, we assume individual 2 died at age 39-0 and individual 3 died at 40-0, which means that both individuals contribute 1 year to $e_{38}$ and individual 3 contributes 1 to $e_{39}$ (note that we do not adjust for lapsed policies). Thus, $e_{38} = (6 + 12 + 12 + 0)/12 = 2.5$ and $e_{39} = 1$, so $\hat{q}_{38} = 1/2.5 = 0.4$ and $\hat{q}_{39} = 1$.

□

### 2.1.1 Insuring Ages

One way that the problem can be altered is to require the use of *insuring ages*. This means that we modify the birthdates in such a way that the insured is assumed to have joined on his birthday for the purposes of modeling. Thus, an individual who has joined at an age of 27-7 has his or her birthdate shifted such that he joined at an age of exactly 27-0. Sometimes, the age is set to the age at the most recent birthday, and other times it is set based on the nearest birthday. The problem will specify which method to use.

**Example 2.2.** Recalculate $\hat{q}_{38}$ using the data in Example 2.1 assuming the company assigned insuring ages by age last birthday and the actuarial exposure method.

**Answer** Individual 1 is now assumed to have a birthdate of 1/1/74, thus entering at 29-0. Individual 2, 3, and 4 similarly have birthdates of 7/1/73, 5/1/74, and 8/1/75, respectively. Thus, we have new termination ages of 37-10, 38-5, 38-7, and 36-11, for the individuals. Now both deaths occur between ages 38 and 39. $e_{38} = (12 + 12)/12 = 2$. Finally, we have $\hat{q}_{38} = 2/2 = 1$.

$\square$

### 2.1.2 Anniversary-to-Anniversary

One additional possible modification to the problem is the use of *anniversary-to-anniversary* method for calculating ages joined/left. (The previously mentioned methods are called *calendar-based* or *date-to-date*.) Under this method, all subjects effectively have their join date shifted forward by one year on the first policy anniversary after the start date of the study. When combined with insuring ages, policyholders are assumed to have an integral age at their now-shifted join date. Any insureds that have not terminated at the study's end have their last observation date shifted earlier to the last anniversary prior to the study end date. Death and surrender ages are still governed by the rules of either actuarial exposure or exact exposure.

**Example 2.3.** Recalculate $\hat{q}_{38}$ using the data in Example 2.1 assuming the company assigned insuring ages by age last birthday, an anniversary-to-anniversary study, and the actuarial exposure method.

**Answer** Individual 1 is now assumed to have a birthdate of 1/1/74 (with insuring ages) and a start date of 1/1/04. Therefore, he joined at age 30-0 and still died at age 37-10. Individual 2 similarly joined at age 30-0 and died at age 38-5, and so on. All we have done is shifted the start ages for everyone in light of the anniversary-to-anniversary study. If anyone was still around at the time the study terminates, we would have had to adjust their ages there as well. Since none of the adjustments made affect the exposures at age 38, we still have the previous answer of $\hat{q}_{38} = 1$.

$\square$

## 2.2 Interval Data

With large studies, it is more common to receive observations in the form of interval data. Interval data loses the granularity of the exact age at which events occur, and instead provide an aggregation of the number of individuals per event per age bin. For an interval from age $c_j$ to age $c_{j+1}$, denote $P_j$ to be the number of individuals being observed at age $c_j$, and let $n_j, d_j$, and $w_j$ be the number entering, dying and leaving during the interval. $P_j$ is almost the exposure for the interval, but we have to modify it to account for people leaving and entering during the interval. We again have two methods of calculating exposure.

1. **Exact exposure**
   We assume within-interval events happen uniformly occur through the entire interval. It is natural to assume that on average, the events occur at the midway point. Thus, we get

$$e_j = P_j + \frac{n_j - d_j - w_j}{2}$$

2. **Actuarial exposure**
   We further assume that deaths do not reduce exposure.

$$e_j = P_j + \frac{n_j - w_j}{2}$$

The Kaplan-Meier estimator also gives the same formula as the actuarial exposure.

Note that in general, $P_{j+1} \neq P_j + n_j - d_j - w_j$ since there are events that occur at precisely the endpoint $c_{j+1}$.

**Example 2.4.** Consider the following data for a study.

| Age | $P_j$ | $n_j$ | $d_j$ | $w_j$ | $P_{j+1}$ |
|-----|-------|-------|-------|-------|-----------|
| 37  | 0     | 3     | 0     | 1     | 5         |
| 38  | 5     | 1     | 1     | 1     | 3         |
| 39  | 3     | 0     | 2     | 0     | 0         |

Calculate the probabilities of withdrawal $\hat{q}_j^w$ at the three ages using exact exposure.

**Answer**

$$e_{37} = 0 + \frac{3 - 0 - 1}{2} = 1$$
$$e_{38} = 5 + \frac{1 - 1 - 1}{2} = 4.5$$
$$e_{39} = 3 + \frac{0 - 2 - 0}{2} = 2$$
$$\hat{q}_{37}^w = 1 - \exp(-w_{37}/e_{37})$$
$$= 1 - \exp(-1/1)$$
$$= 0.6321$$
$$\hat{q}_{38}^w = 1 - \exp(-1/4.5)$$
$$= 0.1993$$
$$\hat{q}_{39}^w = 1 - \exp(-0/2)$$
$$= 0$$

$\square$

## 2.3 Parametric Estimation (and Variances)

Without chugging through the algebra, it turns out the maximum likelihood estimate of $q_j$ is the exact exposure estimate for individual data, if we assume that the survival function decreases exponentially within each interval $[c_j, c_{j+1})$. Namely,

$$\hat{q}_{MLE,j} = 1 - \exp\left(-\frac{(c_{j+1} - c_j)d_j}{e_j}\right)$$

This estimator (MLE and exact exposure estimator) has the variance estimated by

$$\widehat{Var}(\hat{q}_{MLE,j}) = (1 - \hat{q}_{MLE,j})^2 (c_{j+1} - c_j)^2 \frac{d_j}{e_j^2}$$

When dealing with actuarial exposure, we might assume that $d_j$ is binomially distributed with parameters $m = e_j/(c_{j+1} - c_j)$ and $q = \hat{q}_j$. This then leads to the following estimate of the variance of the actuarial exposure estimator:

$$\widehat{Var}(\hat{q}_j) = \frac{\hat{q}_j(1 - \hat{q}_j)}{e_j/(c_{j+1} - c_j)}$$

# 3 J.1 and J.2 (Prerequisite: Chapter 14)

In addition to the standard inverse transform sampling (also known as inversion) method we have learned in Chapter 14, there are some additional ways to sample values from special distributions. In general, these methods are very much rooted in inverse transform sampling, so there is very little added computation complexity. We will now cover these methods.

## 3.1 Discrete Mixtures

Consider a random variable $Y$ which is a mixture of random variables $X_1, \ldots, X_k$, namely its CDF $F_Y(y)$ can be expressed as:
$$F_Y(y) = p_1 F_{X_1}(y) + \cdots + p_k F_{X_k}(y)$$
where $p_1, \ldots, p_k$ are all between 0 and 1 and $\sum p_i = 1$. We can sample a value $y$ from the distribution of $Y$ in the following way.

1. Define a random variable $X$ such that $P(X = k) = p_k$. Simulate a value $x_k$ from the distribution of $X$ via inversion.

2. Simulate a value $y$ from $F_{x_k}(y)$ via inversion.

**Example 3.1.** Consider a random variable $Y$ with CDF $F_Y(y) = 0.7 F_{X_1}(y) + 0.3 F_{X_2}(y)$ where $X_1$ is gamma with $\alpha = 2$ and $\theta = 1.7$ and $X_2$ is exponential with mean 3. Simulate a value from the distribution of $Y$ using the uniform values below:
$$0.85, 0.56$$

**Answer** First, we define $X$ such that $X = 1$ with probability 0.7 and $X = 2$ with probability 0.3. Since our first uniform value of 0.85 is greater than 0.7, we simulate $X = 2$. Thus, we use the second uniform value of 0.56 to simulate from the given exponential distribution.

$$0.56 = 1 - e^{-x/3}$$
$$x = 2.4629$$

$\square$

## 3.2 Normal and Lognormal Distributions

We have also shown how you can simulate from both of these distributions via inversion, but it turns out that the samples may be inexact far away from the mean of the distributions. There are two additional methods which can be used instead. These two methods both generate pairs of independent values from the standard normal distribution at a time.

We will illustrate how to use them for the standard normal distribution, a method which can then be essentially carried over to the normal and lognormal distributions via the standard transformation methods. In general, the polar method outperforms the other methods.

### 3.2.1 Box-Muller Transformation

1. Take as input 2 random uniform values $u_1$ and $u_2$.

2. Compute 2 independent standard normal values given by:

$$z_1 = \sqrt{-2\ln(u_1)} \cos(2\pi u_2)$$
$$z_2 = \sqrt{-2\ln(u_1)} \sin(2\pi u_2)$$

### 3.2.2   Polar Method

1. Take as input 2 random uniform values $u_1$ and $u_2$.

2. Calculate the following values:

$$x_1 = 2u_1 - 1$$
$$x_2 = 2u_2 - 1$$
$$w = x_1^2 + x_2^2$$

3. Repeat steps 1 and 2 if $w \geq 1$. Otherwise let

$$y = \sqrt{-\frac{2\ln(w)}{w}}$$

4. Compute 2 independent standard normal values given by: $z_1 = x_1 y$ and $z_2 = x_2 y$.

### 3.2.3   Example

**Example 3.2.** Simulate 2 values from the lognormal distribution with parameters $\mu = 5$ and $\sigma^2 = 10$ using as many of the following uniform values as needed:

$$0.21, 0.04, 0.19, 0.74, 0.91, 0.27$$

Use both (a) the Box-Muller method and (b) the polar method.

**Answer** (a) We start by simulating 2 values from the standard normal distribution. Starting with $u_1 = 0.21$ and $u_2 = 0.04$, we get $z_1 = \sqrt{-2\ln(0.21)}\cos(2\pi 0.04) = 1.7112$ and $z_2 = 0.4394$. We now want to turn these back into values $v_1$ and $v_2$ from the lognormal distribution. Since $z = (\ln v - \mu)/\sigma$ from our formula sheet, $v_1 = 33,232.70$ and $v_2 = 595.50$.

(b) Again, we start by simulating 2 values from the standard normal distribution. We begin with $u_1 = 0.21$ and $u_2 = 0.04$. Then $x_1 = 2(0.21) - 1 = -0.58$ and $x_2 = 2(0.04) - 1 = -0.92$, so that $w = 1.1828$. Since this value is greater than 1, we proceed to use the next two uniform values.

Doing so, we start with $u_1 = 0.19$ and $u_2 = 0.74$, which leads to $x_1 = -0.62$, $x_2 = 0.48$, $w = 0.6148$, $y = 1.2580$, and finally our standard normal values of $z_1 = -0.7800$ and $z_2 = 0.6038$. Finally, we get $v_1 = 12.60$ and $v_2 = 1001.63$.

$\square$

## 3.3   Simulation from the $(a, b, 0)$ Class

Chapter 14 from the manual already goes over how to simulate from any general discrete distributions via inversion; the new syllabus additionally requires knowledge of how to simulate using stochastic processes. Discrete distributions are normally used to model the number of events that occur over a time interval. Intuitively, simulating via a stochastic process involves simulating the time between events continuously until we have total time elapsed to be greater than 1. Specifically, the steps are:

1. Simulate a series of event intervals $s_k$ from an exponential distribution with mean $1/\lambda_k$ using inverse transform sampling.

2. Stop when the total time $t_T = \sum_{k=0}^{T} s_k > 1$. The simulated value is then $T$.

It is important to note that the parameter $\lambda_k$ varies depending on $k$ and the specific distribution of the $(a, b, 0)$ class. In general, $\lambda_k = c + dk$. Below are the formulas for $c$ and $d$ for each distribution:

| Distribution | $c$ | $d$ |
|---|---|---|
| Poisson($\lambda$) | $\lambda$ | $0$ |
| Binomial($m, q$) | $-md$ | $\ln(1-q)$ |
| Negative Binomial($r, \beta$) | $rd$ | $\ln(1+\beta)$ |

Be careful that for a binomial distribution, it is mathematically possible to still obtain $t_T < 1$ for $T = m$. If this is the case, we would stop and set our simulated value to be $m$.

**Example 3.3.** Simulate a value from a negative binomial distribution with parameters $r = 2$ and $\beta = 1.5$ using a stochastic process and the following values from a uniform distribution:

$$0.15, 0.58, 0.93$$

**Answer** We start by writing out the form of $\lambda_k$. $d = \ln(1+1.5) = 0.9163$ and $c = 2(0.9163) = 1.8326$, so $\lambda_k = 1.8326 + 0.9163k$. The inverse CDF of an exponential distribution parameterized by $\lambda_k$ is $F^{-1}(u) = -\ln(1-u)/\lambda_k$. Thus, we simulate as follows:

$$s_0 = -\frac{\ln(1-0.15)}{1.8326} = 0.0887$$

$$s_1 = -\frac{\ln(1-0.58)}{1.8326 + 0.9163} = 0.3156$$

$$s_2 = -\frac{\ln(1-0.93)}{1.8326 + 0.9163(2)} = 0.7255$$

Note that $t_1 = s_0 + s_1 < 1$ but $t_2 = t_1 + s_2 > 1$. Therefore, we simulate a value of 1.

$\square$

## 3.4 Simulating Multiple Decrements

Death is one possible way of terminating an insurance policy, and the number of deaths is assumed to follow a binomial distribution. What if we also consider other decrements (or even non-terminating outcomes)? The resulting model for consideration is a multinomial distribution.

There are two types of problems that can be asked on this topic. The first involves simulating the number of policyholders having each outcome (a macroscopic question) and the second requires simulating the outcome for a single policyholder. We will show how to solve these problems via two examples.

### 3.4.1 Simulating a Count for Each Outcome

In practice, most problems will simply involve a series of simulations from binomial distributions. (Unless specified otherwise, you should be able to use inversion to sample from the binomial distributions.)

**Example 3.4.** Consider a policy type that has the following probabilities for each insured in the next year: 0.02 (death), 0.05 (lapse), 0.05 (disability), and healthy otherwise. For a study of 100 participants with this policy, simulate the counts of each outcome in the next year using the following numbers from a uniform distribution:

$$0.73, 0.09, 0.12$$

**Answer** The number of deaths follows a binomial distribution with $m = 100$ and $q = 0.02$. Using inversion, we get the number of deaths corresponding to a uniform value of 0.73 to be 3. Next, we simulate the number of lapses, which, conditional on the number of deaths being equal to 3, follows a binomial distribution with $m = 100 - 3$ and $q = 0.05/(1 - 0.02)$. A uniform value of 0.09 from this distribution gives us a total of 2 lapses. Then, finally, we simulate the number of disabled individuals, which conditionally follows a binomial distribution with $m = 100 - 3 - 2$ and $q = 0.05/(1 - 0.05 - 0.02)$. Using a uniform value of 0.12 gives us a total of 3 disabled individuals. All other insureds are healthy.

$\square$

The trick, as seen above, is to reduce $m$ by the simulated value at the previous step, and to update $q$ accordingly.

## 3.5 Simulating Individual Outcomes

This type of question simply requires you to simulate from a categorical variable representing outcomes. The example below should seem fairly simple.

**Example 3.5.** Consider the same setup as in the previous example over a two-year period. Suppose that the probability of death in the first year is equal to the probability of dying in the second year conditional on enrollment in the policy at the start of the second year. Additionally, if the individual is disabled during the first year, then he has an increased probability of death of 0.1 during the second year. Using a uniform value of 0.37, simulate the result of one policy, given the ordering of outcomes given below:

1. Death in year 1.

2. Lapse in year 1.

3. Disabled in year 1, survives year 2.

4. Disabled in year 1, death in year 2.

5. Healthy and insured after year 1, survives year 2.

6. Healthy and insured after year 1, death in year 2.

**Answer** We need to calculate the probabilities corresponding to each outcome and give the associated cumulative probabilities, using the ordering of outcomes provided.

| Outcome | Probability | Cumulative Probability |
|---|---|---|
| Death in year 1 | 0.02 | 0.02 |
| Lapse in year 1 | 0.05 | 0.07 |
| Disabled in year 1, survives year 2 | $0.05(1 - 0.1) = 0.045$ | 0.115 |
| Disabled in year 1, death in year 2 | $0.05(0.1) = 0.005$ | 0.12 |
| Healthy and insured after year 1, survives year 2 | $(1 - 0.02 - 0.07)(1 - 0.02) = 0.8722$ | 0.9922 |
| Healthy and insured after year 1, death in year 2 | $(1 - 0.02 - 0.07)(1 - 0.02) = 0.0178$ | 1 |

Our uniform value of 0.38 then corresponds to being healthy after year 1 and surviving year 2.

□