In the frequentist paradigm, the probability of an event is based on its relative frequency. All prior and/or collateral information is ignored. Proponents of the frequentist paradigm view it as being objective, because all attention is devoted to the observations (data). Some of the key constructs of the frequentist paradigm are the Neymann-Pearson Lemma, tests of statistical hypotheses, confidence intervals, and unbiased estimation.

In the Bayesian paradigm, probability is treated as a rational measure of belief. Thus the Bayesian paradigm is based on personal or subjective probabilities and involves the use of Bayes' Theorem. Prior and/or collateral information is incorporated explicitly into the model via the prior distribution and the likelihood. Some of the key constructs of the Bayesian paradigm, in addition to Bayes' Theorem, are conditional probabilities, prior distributions, predictive distributions, and (posterior) odds ratios.

6.7 Simulating Predictive Distributions

6.7.1 Formulating the Problem

In property and casualty insurance, as well as in health insurance, the actuary is often asked to predict the amount of insured losses during the next period of observation, such as a calendar year. In doing so, the actuary frequently has the results observed for a number of prior periods. Then if X_i represents the loss amount during the i^{th} period of observation, the problem may be considered to be the estimation of the quantity

$$Pr\left[X_{m+1} \le x_{m+1} \mid X_1, X_2, \ldots, X_m\right],$$

where $x_{m+1} \ge 0$. This is the conditional probability distribution of the insured losses incurred during period $m+1$, given the results of periods 1 through m. Such a probability distribution is usually called a *predictive distribution.*

One way of approaching this problem is to first determine the distribution of the frequency of loss (i.e., the number of insurance claims) and then to determine the distribution of the severity or amount of each individual claim. This is referred to as a *two-stage model*, and is the type of problem considered in Chapter 9 of Herzog [29].

Herzog assumes (1) that the random variable representing the number of claims, N_i, during the i^{th} period of observation has a Poisson distribution with parameter (mean) Λ, and (2) that the parameter Λ, in turn, has a gamma distribution. A result of these assumptions, as shown on page 158 of Herzog, is that the predictive distribution of the frequency of claim is a negative binomial distribution.

Moreover, Herzog assumes (1) that the random variable representing the amount of each individual claim has an exponential distribution with parameter (mean) Δ, and (2) that the parameter Δ, in turn, has an inverse gamma distribution. A result of these assumptions, as shown on pages 161 and 162 of Herzog, is that the predictive distribution of the amount of each claim is a Pareto distribution.

We next show how to use pseudo-random numbers and quasi-random numbers to simulate such distributions.

6.7.2 Solving the Problem via a Pseudo-random Number Generator

6.7.2.1 Frequency Component

We assume that the probability of observing n_{m+1} claims during period $m+1$ is given by the negative binomial distribution as

$$Pr[N_{m+1} = n_{m+1}] \quad = \quad \binom{2+n_{m+1}}{n_{m+1}} (.50)^{n_{m+1}} (.50)^3,$$

for $n_{m+1} = 0, 1, \ldots$. We employ a psuedo-random number generator in conjunction with the algorithm for the Modified Table-Look-Up Approach to the negative binomial distribution, given in Section 3.2.5.1, to simulate 10,000 trials of the number of claims. The results are summarized in Table 6.4 below.

For the 10,000 trials simulated here we have observed a total of 30,278 claims, which is slightly more than the 3 claims per trial that are expected. (See the discussion in Section 3.2.5 for more details.)

TABLE 6.4

Frequency Component Constructed Using Pseudo-random Number Generator

Number of Claims	Frequency of Occurrence
0	1,234
1	1,852
2	1,910
3	1,548
4	1,110
5	847
6	572
7	356
8	237
9	128
10	93
11	47
12	33
13	14
14	7
15	4
16	6
17	0
18	2
Total	**30,278**

6.7.2.2 Severity Component

For each of the 30,278 individual claims of the previous section, we need to simulate an individual loss (or claim) amount. We do this by using a pseudo-random number generator to produce uniform random numbers over [0,1), in conjunction with the inversion scheme of Section 3.1.6 applied to the Pareto probability distribution function given by

$$F(x) = \begin{cases} 0 & x \le 0 \\ 1 - \frac{\beta^{\alpha}}{(x+\beta)^{\alpha}} & x > 0 \end{cases},$$

where $\alpha = 20$ and $\beta = 2{,}000{,}000$. In particular, if U is the result of simulating a uniform random variable over [0,1), then the corresponding Pareto random variable is $\frac{\beta}{(1-U)^{1/\alpha}} - \beta$. The results are summarized by the loss severity distribution for which various percentiles are displayed in Table 6.5.

TABLE 6.5

Loss Severity Distribution Constructed Using Pseudo-random Number Generator

Percentile	Point
0	$ 4
10	10,930
25	29,148
50	70,435
75	144,126
90	244,628
100	1,447,454

6.7.2.3 Loss Amounts

Finally, we employ the results of Sections 6.7.2.1 and 6.7.2.2 to produce the distribution of individual loss amounts summarized in Table 6.6. To illustrate the process, if an individual trial resulted in two claims, then we drew two values from the loss severity distribution.

TABLE 6.6

Distribution of Loss Amounts Using Pseudo-random Numbers

Percentile	Point
0	$ 0
10	0
25	71,590
50	229,179
75	471,889
90	753,763
100	2,738,627

6.7.3 Solving the Problem Using Quasi-random Numbers

Because we do not know in advance how many quasi-random numbers we need as input to the algorithm employed to simulate the negative binomial distribution, we do not employ a quasi-Monte Carlo scheme to simulate the number of claims. If we attempted to do so, we would end up with a biased result. However, we can use a quasi-Monte Carlo scheme to simulate the severity portion of the problem, the loss amounts on the 30,278 claims resulting from the first stage of our model.

Our approach is to employ the Neiderreiter sequence

$$\left\{\frac{1}{60{,}556}, \frac{3}{60{,}556}, \ldots, \frac{60{,}555}{60{,}556}\right\},$$

where $60{,}556 = 2\times 30{,}278,$ to obtain values from the above Pareto distribution. This gives us an empirical loss severity distribution consisting of 30,278 individual claim amounts in ascending order. The results are summarized in Table 6.7.

TABLE 6.7

Loss Severity Distribution Using Quasi-random Numbers

Percentile	Point
0	$ 2
10	10,563
25	28,974
50	70,526
75	143,540
90	244,011
100	1,468,469

We then employed a pseudo-random number generator to obtain a "random" permutation of the integers from 1 to 30,278, in order to "randomly" re-order (or "shuffle") these loss amounts. The loss amounts are then assigned to an individual trial to produce the distribution of loss amounts summarized in Table 6.8 below.

TABLE 6.8

Distribution of Loss Amounts Using Quasi-random Numbers to Generate the Severity of Loss

Percentile	Point
0	$ 0
10	0
25	67,749
50	227,986
75	467,540
90	753,417
100	2,969,518

Because the quasi-random numbers were “superior” in our previous comparisons, we suspect that the results of Table 6.8 are “superior” to those of Table 6.6.