

I. Review of Probability Theory

Our principal goal, in this section, is to quickly review the basic ideas and concepts of probability theory. For additional background and examples, please consult the text used for your prerequisite probability class and/or the following books:

- *Introduction to Probability Models* by Sheldon M. Ross, Academic Press (Chapters 1 to 3)
- *Probability and Statistics with Reliability, Queueing, and Computer Science Applications* by Kishor S. Trivedi, Prentice-Hall (Chapters 1 to 5)
- *Probability, Statistics, and Stochastic Processes* by Peter Olofsson, John Wiley (Chapters 1 to 4)

1 What is Stochastic Modeling?

Stochastic modeling concerns the use of probability to model real-world situations in which uncertainty is present. Since uncertainty is pervasive, this means that the tools of this course can potentially prove useful in almost all facets of one's professional life (and sometimes even in one's personal life):

- Gambling
- Personal Finances
- Disease Treatment Options
- Economic Forecasting
- Product Demand
- Call Center Provisioning
- Product Reliability and Warranty Analysis, etc.

The use of a stochastic model does not imply that the modeler fundamentally believes that the system under consideration behaves "randomly". (For example, the behavior of an individual may appear "random". But an interview with that individual may reveal a set of preferences under which that person's behavior is then revealed as totally predictable.) Use of a stochastic model reflects only a pragmatic decision on the part of the modeler that such a model represents the best currently available description of the phenomenon under consideration, given the data that is available and the universe of models known to the modeler.

2 The Basic Steps of Stochastic Modeling

The essential steps in building stochastic models are:

- i) Identifying the sample space;
- ii) Assigning probabilities to the elements of the sample space;
- iii) Identifying the events of interest;
- iv) Computing the desired probabilities.

3 Elements of Basic Probability

The set of all possible outcomes that are present in a stochastic model is called the *sample space*, typically denoted as Ω . A typical element of Ω (or “outcome”) is often denoted as $\omega \in \Omega$.

Example 1.3.1: In tossing a coin, $\Omega = \{H, T\}$ is one possible choice for the sample space. Alternatively, we could have chosen $\Omega = \{0, 1\}$ (0=“tails”; 1=“heads”).

Example 1.3.2: When shuffling a deck of 52 cards, a typical outcome can be viewed as a permutation σ of the 52 integers $\{1, \dots, 52\}$. The sample space Ω is the set of 52! permutations on the integers $\{1, \dots, 52\}$.

Example 1.3.3: In testing product defects coming off a production line, one possible choice of sample space is $\Omega = \{(i_1, i_2, \dots) : i_j \in \{0, 1\} \text{ for } j \geq 1\}$, where $i_j = 1$ if the j 'th item sampled is defective and $i_j = 0$ otherwise.

Example 1.3.4: Suppose one is studying the lifetime of an individual. Here, we could set $\Omega = [0, \infty)$ (or use $\Omega = [0, 130]$, since 130 is an upper bound on human lifetimes).

An *event* is a subset of Ω . A *probability* P assigns numbers between 0 and 1 to events. The probability $P(A)$ of an event A is interpreted as the “likelihood” of the event A (under the probability model P that has been chosen). In order that P assigns probabilities in a logically consistent manner, P must have the following properties:

- i) $P(\Omega) = 1$.
- ii) For any sequence of mutually exclusive events A_1, A_2, \dots

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Example 1.3.1 (continued): If the coin is “fair” (or “unbiased”), $P(H) = P(T) = 1/2$. Otherwise, if $P(H) \neq 1/2$, the coin is said to be *biased*.

Example 1.3.2 (continued): If the deck has been properly shuffled, all $52!$ permutations should be equally likely, so $P(\sigma) = 1/52!$ for each permutation $\sigma \in \Omega$.

Example 1.3.3 (continued): One possible probability assignment is the one for which

$$P(A_{k,n}) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $p \in [0, 1]$, where $A_{k,n}$ is the event corresponding to the set of all outcomes in which exactly k of the first n products tested are defective.

Example 1.3.4 (continued): Here, the typical form of a probability P is one for which there exists a non-negative function f (integrating to 1) for which

$$P(A) = \int_A f(x) dx$$

The key mathematical idea in building stochastic models is the concept of *conditional probability*. The conditional probability of A , given that B has occurred, is denoted as $P(A|B)$ and is defined as

$$P(A|B) \triangleq \frac{P(A \cap B)}{P(B)}$$

where $P(B) > 0$.

Example 1.3.5: A family has two children. The sample space for the children's genders is $\Omega = \{(G, G), (G, B), (B, G), (B, B)\}$. Given that the likelihood of giving birth to a girl is almost the same as the likelihood of giving birth to a boy, the natural probability assignment is $1/4$ to all four outcomes. What is the probability that both children are girls, given that at least one is a girl?

Here, $A = \{(G, G)\}$ and $B = \{(G, G), (G, B), (B, G)\}$. So,

$$P(A|B) = \frac{1/4}{3/4} = \frac{1}{3}.$$

In building models, we often wish to involve the concept of "causality". In a stochastic setting, this means that A becomes more likely in the presence of B having occurred, so that $P(A|B) > P(A)$. If B does not influence A , then $P(A|B) = P(A)$ and the two events are said to be *independent*. Independence is equivalent to requiring that

$$P(A \cap B) = P(A)P(B).$$

More generally, we say that the n events A_1, A_2, \dots, A_n are independent if

$$P(B_1 \cap B_2 \cap \dots \cap B_n) = P(B_1)P(B_2) \cdots P(B_n),$$

where (for each i), B_i equals either A_i or A_i^c .

Example 1.3.3 (continued): If A_i is the event that the i 'th item tested is defective, then A_1, A_2, \dots, A_n are independent events under the probability P described earlier.

Because stochastic models are often specified in terms of conditional probabilities, the following proposition for computing the probability of an event A is frequently useful.

Proposition 1.3.1: If B_1, B_2, \dots, B_n are mutually exclusive events, then

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i).$$

Example 1.3.6: Suppose we receive 500 items from Factory A and 1000 items from Factory B. Factory A has a defective rate of 1%, whereas Factory B has a defective rate of 0.5%. What is the probability a randomly chosen item is defective?

$$\begin{aligned} P(\text{defective}) &= P(\text{defective}|\text{manufactured at A}) \cdot \frac{500}{1500} \\ &+ P(\text{defective}|\text{manufactured at B}) \cdot \frac{1000}{1500} \\ &= (0.01) \left(\frac{1}{3}\right) + (0.005) \left(\frac{2}{3}\right) = 0.0067. \end{aligned}$$

In many problem settings, we wish to compute $P(A_i|B)$, where the $P(B|A_i)$'s are easily determined from the model specifications. The key to such calculations is *Baye's formula*.

Proposition 1.3.2: Let A_1, A_2, \dots, A_n be mutually exclusive events. Then,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}$$

Example 1.3.7: Suppose that 1% of all children have tuberculosis (TB). When a child who has TB is given what is called the Mantoux test, a positive result occurs 99% of the time. When the child does not have TB, one gets a positive result 1% of the time. Suppose that a child is tested with a positive outcome. What is the likelihood the child actually has TB?

Let A_1 be the event that the child has TB and $A_2 = A_1^c$. If B is the event that the child tests positive, then

$$P(A_1|B) = \frac{(0.99)(0.01)}{(0.99)(0.01) + (0.01)(0.99)} = 0.49$$

So, the likelihood is a bit less than 50%!

4 Random Variables

The systematic use of “random variables” provides a powerful mathematical mechanism for formulating complex stochastic models and performing related calculations. Formally, a *random variable* X (rv) is a function that maps each outcome $\omega \in \Omega$ into a number.

Example 1.3.2 (continued): A typical outcome here is a permutation $\omega = \sigma$. A possible rv of interest is $X(\omega) =$ number of face cards drawn in a hand of 3 cards. Note that

$$P(\{\omega : X(\omega) = 0\}) = \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50}.$$

In order to reduce the notational burden, we usually suppress the dependence on ω , and write the above as

$$P(X = 0) = \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50}.$$

Example 1.3.3 (continued): Suppose that $X_i = 1$ if the i 'th item tested is defective and 0 otherwise. For the probability specification given earlier,

$$P(X_i = 1) = p = 1 - P(X_i = 0).$$

A rv X is said to be *discrete* if it can take on only a discrete set of possible values (i.e., the set of possible values is either finite or “countably infinite”). Each of the rv's X given above is discrete. A rv X taking on a continuum of possible values is said to be a *continuous* rv.

Example 1.3.4 (continued): Suppose that $X(\omega) = \omega$, so that X is the lifetime of the individual. Then, X is a continuous rv and

$$P(X \in A) = \int_A f(x) dx.$$

In communicating the probabilistic characteristics of a rv, it is common to speak of the *distribution function* of a rv. In particular, the distribution function of the rv X is the function

$$F(x) = P(X \leq x).$$

Note that it is non-decreasing, with $F(-\infty) = 0$ and $F(\infty) = 1$.

If X is discrete, it is common to work with the *probability mass function* (pmf) given by

$$p(x) = P(X = x).$$

Note that

$$F(x) = \sum_{y \leq x} p(y).$$

On the other hand, if X is continuous, one often works with the *probability density function* (pdf) $f(x)$. By analogy with the discrete case, F is related to the density f via

$$F(x) = \int_{-\infty}^x f(t) dt,$$

so that

$$f(x) = \frac{d}{dx} F(x).$$

In contrast to the discrete case, the density $f(x)$ can not be directly interpreted as a probability. In particular, $f(x)$ need not be less than 1. But it can be interpreted as a (relative) likelihood. Suppose, for example, that we are interested in the relative likelihood

$$\frac{P(X \in [b, b + h])}{P(X \in [a, a + h])}$$

for h small. Note that

$$P(X \in [b, b+h]) = F(b+h) - F(b) \approx h \cdot f(b)$$

when h is small. So,

$$\frac{P(X \in [b, b+h])}{P(X \in [a, a+h])} \approx \frac{h \cdot f(b)}{h \cdot f(a)} \rightarrow \frac{f(b)}{f(a)}$$

as $h \downarrow 0$. (A more careful argument would use l'Hospital's rule.) So, $f(b)/f(a)$ can indeed be interpreted as the relative likelihood that X takes on value b rather than a .

5 Key Random Variables

The following discrete rv's arise in many modeling contexts:

Bernoulli rv A rv X taking on the values 0 and 1 only. The rv X is said to be a Bernoulli rv with parameter p ($\in [0, 1]$) if

$$P(X = 1) = p = 1 - P(X = 0).$$

Binomial (n, p) rv A rv X is said to be a Binomial (n, p) rv if it takes on only the values $0, 1, \dots, n$ and

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Geometric (p) rv The rv X taking values in $\{0, 1, 2, \dots\}$ is said to be geometric with parameter p if

$$P(X = k) = p(1-p)^k.$$

Poisson (λ) rv The rv X taking values in $\{0, 1, 2, \dots\}$ is Poisson with parameter $\lambda \in (0, \infty)$ if

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

The following are some of the key continuous rv's:

Uniform $[a, b]$ rv The rv X is said to be uniform on $[a, b]$ if it has the density

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{else} \end{cases}$$

Exponential (λ) rv The rv X is exponentially distributed with parameter $\lambda > 0$ if it has the density

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & \text{else} \end{cases}$$

Weibull (λ, α) **rv** A Weibull (λ, α) rv has the density

$$f(x) = \begin{cases} \alpha \lambda^\alpha x^{\alpha-1} \exp(-(\lambda x)^\alpha), & x \geq 0, \\ 0, & \text{else} \end{cases}$$

The parameter λ is called the “scale” parameter and α is called the “shape” parameter.

Gamma (λ, α) **rv** The rv X is said to be gamma distributed with scale parameter λ and shape parameter α if it has the density

$$f(x) = \begin{cases} \lambda(\lambda x)^{\alpha-1} \exp(-(\lambda x)), & x \geq 0, \\ 0, & \text{else} \end{cases}$$

The special case where α is a positive integer is known as an Erlang rv; the integer α is called the “number of stages” of the rv.

Normal (μ, σ^2) **rv** The rv X is said to be normally distributed with mean μ and variance σ^2 if it has the density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Such a rv is also often referred to as a Gaussian rv (in honor of the great mathematician Carl Friedrich Gauss), and is typically denoted as a $N(\mu, \sigma^2)$ rv.

6 Expectation of a Random Variable

Suppose that one needs to communicate the key probability characteristics of a rv X . One possible approach is to communicate the entire probability mass function or density function.

But a more succinct summary quantity is often desirable. Here are some of the common choices:

mode of the rv X This is the value x^* that maximizes $f(x)$ (when X is continuous) and $p(x)$ (when X is discrete). For example, the mode of an $\text{Exp}(\lambda)$ rv is 0.

median of the rv X This is the value m with the property that

$$P(X \leq m) = \frac{1}{2} = P(X \geq m),$$

when X is continuous. Note that when X is $\text{Exp}(\lambda)$,

$$m = \frac{1}{\lambda} \log(2).$$

Note : For the purposes of this course, all logarithms are “natural” (i.e., to the base e).

expectation of the rv X The expectation of a discrete rv X , denoted $E(X)$, is the quantity

$$E(X) = \sum_x xp(x).$$

The expectation of a continuous rv X , also denoted $E(X)$, is the quantity

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx.$$

For an $\text{Exp}(\lambda)$ rv, the expectation is $1/\lambda$.

It is often the case that one must determine how the probability characteristics of a rv X are modified by a transformation to the rv $g(X)$.

Example 1.6.1 Suppose that U is uniformly distributed on $[0, 1]$, so that its density is

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1, \\ 0, & \text{else} \end{cases}$$

We are interested in U^3 . To compute $E(U^3)$, set $X = U^3$. Based on the above discussion, we need to calculate the density of U^3 . We do this by first computing the distribution function of $X = U^3$. Observe that

$$P(X \leq x) = P(U^3 \leq x) = P(U \leq x^{1/3}) = x^{1/3}$$

for $0 \leq x \leq 1$. So, the density is

$$\frac{d}{dx}P(X \leq x) = \begin{cases} \frac{1}{3}x^{-2/3}, & 0 \leq x \leq 1, \\ 0, & \text{else} \end{cases}$$

Hence,

$$E(U^3) = \int_0^1 x \cdot \frac{1}{3}x^{-2/3} dx = \frac{1}{3} \int_0^1 x^{1/3} dx = \frac{1}{4}.$$

But the next result establishes that there is a (much!) faster way of doing such calculations.

Theorem 1.6.1 If X is a discrete rv with pmf $p(x)$, then

$$E(g(X)) = \sum_x g(x)p(x).$$

If X is a continuous rv with pdf $f(x)$, then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x) dx.$$

Example 1.6.1 (continued) According to Theorem 1.6.1, we can compute $E(U^3)$ as follows:

$$E(U^3) = \int_0^1 x^3 f(x) dx = \int_0^1 x^3 dx = \frac{1}{4}.$$

Much faster!

One also frequently needs succinct summary quantities to characterize the amount of “random fluctuation” in X . Here are several common measures:

variance of the rv X The variance of the rv X , denoted $\text{var}(X)$, is the quantity

$$\text{var}(X) = \text{E}[(X - \text{E}(X))^2].$$

According to Theorem 1.6.1,

$$\begin{aligned}\text{var}(X) &= \int_{-\infty}^{\infty} (x - \text{E}(X))^2 f(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f(x) dx - (\text{E}(X))^2 \\ &= \text{E}(X^2) - (\text{E}(X))^2\end{aligned}$$

when X is continuous ; the same formula holds in the discrete case. One problem with the variance is that it does not “scale” properly. For example, suppose that X is the height of a randomly chosen individual, measured in feet. If one now changes the unit of measure to inches, this multiplies the variance by a factor of 144!

standard deviation of the rv X To remedy this scaling problem in the variance, a more meaningful measure of variability is the standard deviation, denoted $\sigma(X)$:

$$\sigma(X) = \sqrt{\text{var}(X)}.$$

The standard deviation of an $\text{Exp}(\lambda)$ rv is easily seen to be $1/\lambda$. Note that the standard deviation scales in the same way as does the expectation.

inter-quartile range For X a continuous rv, the inter-quartile range is just

$$F^{-1}\left(\frac{3}{4}\right) - F^{-1}\left(\frac{1}{4}\right),$$

where F^{-1} is the inverse function to the distribution function F . For an $\text{Exp}(\lambda)$ rv, the inter-quartile range is just $\frac{1}{\lambda} \log 3$.

7 Jointly Distributed Random Variables

In most modeling contexts, we will need to understand the interactions between multiple rv’s. This requires the notion of joint distributions of rv’s. We illustrate with the case of two rv’s; everything generalizes in a straightforward way to a collection of n rv’s.

Given a pair of rv’s X and Y , the joint cumulative distribution function F is given by

$$F(x, y) = P(X \leq x, Y \leq y).$$

Note that the distribution function of either X or Y can be easily recovered from the joint distribution:

$$\begin{aligned}P(X \leq x) &= F(x, \infty), \\ P(Y \leq Y) &= F(\infty, y).\end{aligned}$$

If both X and Y are discrete, there is a joint probability mass function (pmf) p such that

$$P(X = x, Y = y) = p(x, y)$$

and

$$F(x, y) = \sum_{x' \leq x, y' \leq y} p(x', y').$$

On the other hand, if X and Y are both continuous, there typically exists a joint probability density function (pdf) f such that

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x', y') dx' dy'$$

and the joint pdf can be recovered from the joint distribution function by the formula

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y).$$

Theorem 1.6.1 goes over to the case where g depends on more than one random variable.

Theorem 1.7.1 If X and Y are discrete rv's with joint pmf $p(x, y)$, then

$$E(g(X, Y)) = \sum_{x, y} g(x, y)p(x, y).$$

If X and Y are both continuous rv's with joint pdf $f(x, y)$, then

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y) dx dy.$$

In particular, if $g(x, y) = x + y$, then

$$\begin{aligned} E(X + Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y)f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f(x, y) dy dx + \int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy \\ &= E(X) + E(Y), \end{aligned}$$

where $f_X(x)$ and $f_Y(y)$ are the (marginal) pdf's of X and Y , respectively. So, the expectation operation is a linear operation: The expectation of the sum is the sum of the expectations.

Example 1.7.1 Suppose one is collecting coupons to obtain a retail discount. There are r different types of coupons. We assume that the probability that a coupon is of type i is $1/r$. How many different types of coupons will we find, in expectation, if we collect n coupons?

Let $X_i = 1$ if a type- i coupon is present in the collection of n coupons and 0 otherwise. We want to compute $E(X_1 + \dots + X_r)$. By linearity of expectation, this equals $E(X_1) + \dots + E(X_r)$. If the identities of the coupons collected are independent rv's, then

$$\begin{aligned} E(X_i) &= 0 \cdot P(X_i = 0) + 1 \cdot P(X_i = 1) \\ &= P(X_i = 1) \\ &= 1 - P(X_i = 0) \\ &= 1 - \left(\frac{r-1}{r}\right)^n. \end{aligned}$$

So, the expected number of distinct coupon types is $n(1 - (\frac{r-1}{r})^n)$.

8 Independence of Random Variables

We say that a collection of rv's X_1, X_2, \dots, X_n is *mutually independent* if

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = \prod_{i=1}^n P(X_i \leq x_i)$$

for all x_1, \dots, x_n . If the X_i 's are discrete, this implies that the joint pmf factorizes:

$$p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(X_i = x_i).$$

Similarly, if the X_i 's are continuous, the joint pdf factorizes:

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i).$$

where $f(x_i)$ is the (marginal) density of the rv X_i .

Note that if X and Y are independent continuous rv's, this implies that

$$\begin{aligned} E(g(X)h(Y)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} g(x)f_X(x) dx \cdot \int_{-\infty}^{\infty} h(y)f_Y(y) dy \\ &= E(g(X)) \cdot E(h(Y)); \end{aligned}$$

the same argument works for the discrete rv's.

9 Covariance and Correlation

A succinct summary measure that quantifies the dependence between two rv's X and Y is the *covariance* between X and Y , defined by

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))].$$

It is easily verified that

$$\text{cov}(X, Y) = E(XY) - E(X)E(Y),$$

and that $\text{cov}(X, X) = \text{var}(X)$. Furthermore, covariance is “multi-linear”, in the sense that

$$\text{cov}(X_1 + X_2, Y_1 + Y_2) = \text{cov}(X_1, Y_1) + \text{cov}(X_1, Y_2) + \text{cov}(X_2, Y_1) + \text{cov}(X_2, Y_2).$$

It follows that if $S = X_1 + X_2 + \cdots + X_n$, we can compute $\text{var}(S)$ in terms of the covariances between the X_i 's:

$$\text{var}(S) = \text{cov}(S, S) = \sum_{i=1}^n \text{var}(X_i) + 2 \sum_{i=1}^n \sum_{j=i+1}^n \text{cov}(X_i, X_j).$$

So, if the X_i 's are independent and identically distributed (iid), the covariance between X_i and X_j vanishes (for $i \neq j$) and

$$\text{var}(S) = n\text{var}(X_1).$$

Note that, just as for the variance, the covariance does not “scale” properly as a function of the unit in which X and Y are measured. A unitless measure of dependence is the *correlation* between X and Y , defined by

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}.$$

The quantity $\text{corr}(X, Y)$ always lies in the interval $[-1, 1]$.

10 Sums of Independent Random Variables

Suppose that X_i is the duration of task i . Let $S_n = X_1 + X_2 + \cdots + X_n$ denote the time required to perform the tasks 1 through n (assuming that the tasks are performed sequentially). Assume that the X_i 's are independent continuous rv's having densities f_1, f_2, \dots, f_n . Then,

$$P(S_n \leq x) = \int \cdots \int_{x_1 + \cdots + x_n \leq x} f_1(x_1) f_2(x_2) \cdots f_n(x_n) dx_1 dx_2 \cdots dx_n.$$

In particular, if $n = 2$,

$$\begin{aligned} P(X_1 + X_2 \leq x) &= \iint_{x_1 + x_2 \leq x} f_1(x_1) f_2(x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{x-x_2} f_1(x_1) dx_1 f_2(x_2) dx_2 \\ &= \int_{-\infty}^{\infty} F_1(x - x_1) f_2(x_2) dx_2, \end{aligned}$$

where $F_1(x)$ is the distribution function of X_1 . So the density of $X_1 + X_2$ is

$$\begin{aligned} f(x) &= \int_{-\infty}^{\infty} F_1(x - x_1) f_2(x_2) dx_2 \\ &= \int_{-\infty}^{\infty} f_1(x - x_1) f_2(x_2) dx_2, \end{aligned}$$

An integral of the form

$$\int_{-\infty}^{\infty} h_1(x-y)h_2(y) dy$$

is called a *convolution* integral (specifically, the convolution of the functions h_1 and h_2). So, the density of the sum $X_1 + X_2$ is the convolution of the densities f_1 and f_2 .

Example 1.10.1 Suppose that X_1, X_2, \dots, X_n are iid $\text{Exp}(\lambda)$ rv's. Then, the probability density of $X_1 + X_2$ is

$$\begin{aligned} & \int_{-\infty}^{\infty} f_1(x-y)f_2(y) dy \\ &= \int_0^{\infty} f_1(x-y)\lambda e^{-\lambda y} dy \quad (\text{since } f_2(y) \text{ vanishes for } y < 0) \\ &= \int_0^x \lambda e^{-\lambda(x-y)}\lambda e^{-\lambda y} dy \quad (\text{since } f_1(x-y) \text{ vanishes for } y > x) \\ &= \lambda^2 e^{-\lambda x} \int_0^x dy = \lambda^2 x e^{-\lambda x}, \end{aligned}$$

which is the density of an Erlang rv with scale parameter λ and two stages. More generally, if $f^{(n)}(x)$ is the density of $X_1 + X_2 + \dots + X_n$,

$$f^{(n)}(x) = \int_0^x f^{(n-1)}(x-y)\lambda e^{-\lambda y} dy = \lambda \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x},$$

namely, the density of an Erlang rv with scale parameter λ and n stages.

Typically, the n -fold convolution $f^{(n)}(x)$ corresponding to the density of the sum of $X_1 + \dots + X_n$ can not be computed in closed form.

11 Conditional Distributions

Given a joint pmf for two discrete rv's X and Y , the conditional distribution of X , given $Y = y$, is

$$F(x|y) \triangleq P(X \leq x|Y = y) = \frac{\sum_{x' \leq x} p(x', y)}{\sum_{x'} p(x', y)}$$

and the conditional pmf of X , given $Y = y$, is

$$p(x|y) = \frac{p(x, y)}{p_Y(y)},$$

where $p_Y(y)$ is the (marginal) pmf of Y .

Similarly, if X and Y are continuous jointly distributed rv's with joint density $f(x, y)$, the conditional distribution of X , given $Y = y$, is

$$F(x|y) = \frac{\int_{-\infty}^x f(x', y) dx'}{\int_{-\infty}^{\infty} f(x', y) dx'}$$

and the conditional pdf of X , given $Y = y$, is

$$f(x|y) = \frac{f(x, y)}{f_Y(y)},$$

where $f_Y(y)$ is the (marginal) density of Y .

Note that we can compute the (marginal) density of X via conditioning:

$$f_X(x) = \int_{-\infty}^{\infty} f(x|y)f_Y(y) dy.$$

Furthermore, if $E(X|Y = y)$ is the expectation of the conditional distribution of X , given $Y = y$, given by

$$E(X|Y = y) = \int_{-\infty}^{\infty} xf(x|y) dx,$$

we can compute $E(X)$ by conditioning:

$$E(X) = \int_{-\infty}^{\infty} E(X|Y = y)f_Y(y) dy.$$

12 Limit Theorems and Approximations

As one quickly discovers when doing stochastic modeling, many problems lead to computations that are extremely difficult or impossible to implement as “closed form” calculations. In the setting of such problems, one has alternatives:

- i) Try to develop an approximation to the quantity of interest;
- ii) Use the computer to calculate the quantity of interest.

In the early (pre-computer) days of probability, virtually all such research focused on the first alternative, namely development of good approximations. More recently, computational methods have become interestingly important. Nevertheless, there are many problems that continue to be challenging from a computational viewpoint and for which approximations continue to be the preferred approach.

The use of approximations leads naturally to the question: How does one know that one has developed a good approximation? To answer this question, consider the approximation

$$\left(1 + \frac{x}{n}\right)^n \approx e^x \tag{1.12.1}$$

that one frequently sees in the setting of a calculus course. The above approximation can be justified on the basis of the limit result

$$\left(1 + \frac{x}{n}\right)^n \rightarrow e^x \tag{1.12.2}$$

as $n \rightarrow \infty$. In other words, the limit (1.12.2) makes clear that the approximation (1.12.1) is asymptotically valid as $n \rightarrow \infty$. From a practical standpoint, this means that (1.12.1) is a good approximation when the parameter n is large.

In the world of stochastic modeling, approximations are mathematically justified through the use of limit theorems. This raises the question of what convergence means, when dealing with random quantities. For the purposes of this course, we will need only two convergence concepts.

The first convergence concept that we will need is that of *convergence in distribution*.

Definition 1.12.1 Suppose that $(X_n : 1 \leq n \leq \infty)$ is a sequence of discrete rv's. We say that X_n converges in distribution to X_∞ as $n \rightarrow \infty$ (and write $X_n \Rightarrow X_\infty$ as $n \rightarrow \infty$) if, for each x ,

$$P(X_n = x) \rightarrow P(X_\infty = x)$$

as $n \rightarrow \infty$. Similarly, if $(X_n : 1 \leq n \leq \infty)$ is a sequence of continuous rv's, we say that X_n converges in distribution to X_∞ as $n \rightarrow \infty$ (and write $X_n \Rightarrow X_\infty$ as $n \rightarrow \infty$) if, for each x ,

$$P(X_n \leq x) \rightarrow P(X_\infty \leq x)$$

as $n \rightarrow \infty$.

In other words, $P(X_n \leq x) \approx P(X_\infty \leq x)$ when n is large, so that the distribution of X_n is well-approximated by that of X_∞ when n is large. When such an approximation is valid, we shall write

$$X_n \stackrel{\mathfrak{D}}{\approx} X_\infty.$$

(i.e. X_n "has approximately the same distribution" as X_∞). Given that most probability calculations involve computing distributions, this mode of convergence is especially useful.

We will illustrate the notion of convergence in distribution with the following result. Let Binomial(n, p) be a binomially distributed rv with parameters n and p , and let Poisson(λ) be a Poisson distributed rv with parameter λ .

Proposition 1.12.1 If $np \rightarrow \lambda$ as $n \rightarrow \infty$, then

$$\text{Binomial}(n, p) \Rightarrow \text{Poisson}(\lambda)$$

as $n \rightarrow \infty$.

Proof: Note that

$$\begin{aligned} P(\text{Binomial}(n, p) = k) &= \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k} \\ &\approx \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \left(\frac{n}{n}\right)^k \cdots \left(\frac{n-k+1}{n}\right)^k \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &\rightarrow \frac{\lambda^k}{k!} e^{-\lambda} = P(\text{Poisson}(\lambda) = k). \quad \square \end{aligned}$$

This limit theorem justifies the approximation

$$\text{Binomial}(n, p) \stackrel{\mathfrak{D}}{\approx} \text{Poisson}(np) \tag{1.12.3}$$

when n is large and p is small. A major advantage of the above approximation is that the pmf of the Poisson rv avoids the large factorials that tend to appear in the corresponding pmf of the binomial rv when n is large, thereby avoiding numerical "overflow" problems that can occur when implementing floating-point arithmetic.

Example 1.3.3(continued) Suppose that S_n is the number of defectives found in the first n items. The rv S_n is binomially distributed with parameters n and p . If $n = 100$ and $p = 0.01$, the approximation (1.12.3) establishes that the probability of no defects is

$$P(S_n) = 0 \approx P(\text{Poisson}(1) = 0) = e^{-1}.$$

The second convergence concept that we will use in this class is that of *convergence in probability*.

Definition 1.12.2 Suppose that $(X_n : 1 \leq n \leq \infty)$ is a sequence of rv's. We say that X_n converges in probability to X_∞ (and write $X_n \xrightarrow{P} X_\infty$) if, for each $\epsilon > 0$,

$$P(|X_n - X_\infty| > \epsilon) \rightarrow 0$$

as $n \rightarrow \infty$.

This notion of convergence asserts that with probability approaching to 1, X_n will be close (within ϵ) to X_∞ when n is large.

Proposition 1.12.2 If $X_n \xrightarrow{P} X_\infty$ as $n \rightarrow \infty$, then $X_n \Rightarrow X_\infty$ as $n \rightarrow \infty$.

In other words, convergence in probability implies convergence in distribution. The converse is, however, false.

13 The Law of Large Numbers

One of the two most important results in probability is the law of large numbers (LLN).

Theorem 1.13.1 Suppose that $(X_n : n \geq 1)$ is a sequence of iid rv's. Then,

$$\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow{P} E(X_1)$$

as $n \rightarrow \infty$.

This result is easily to prove when X_i 's have finite variance. The key is the following inequality, called *Markov's inequality*.

Proposition 1.13.1 Suppose that W is a non-negative rv. Then,

$$P(W > w) \leq \frac{1}{w}E(W).$$

Proof: Note that if W is a continuous rv,

$$\begin{aligned} P(W > w) &= \int_w^\infty f(x) dx \\ &\leq \int_w^\infty \left(\frac{x}{w}\right) f(x) dx \quad (\text{since } \frac{x}{w} \geq 1 \text{ when } x \geq w) \\ &\leq \int_0^\infty \left(\frac{x}{w}\right) f(x) dx = \frac{1}{w}E(W). \end{aligned}$$

The proof is similar for discrete rv's. \square

An important special case is called *Chebyshev's inequality*.

Proposition 1.13.2 Suppose that X_i 's are iid with common (finite) variance σ^2 . If $S_n = X_1 + \dots + X_n$, then

$$P\left(\left|\frac{S_n}{n} - E(X_1)\right| > \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Proof: Put $W = (S_n - nE(X_1))^2$ and $w = n^2\epsilon^2$. Note that $E(W) = \text{var}(S_n) = n\sigma^2$, so

$$P\left(\left|\frac{S_n}{n} - E(X_1)\right| > \epsilon\right) = P(W > w) \leq \frac{\sigma^2}{n\epsilon^2}. \quad \square$$

Theorem 1.13.1 is an immediate consequence of Proposition 1.13.2. Let's now apply the LLN.

Example 1.3.3(continued) Note that the proportion of defectives found in the first n item tested is $(X_1 + \dots + X_n)/n$. The LLN asserts that

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{P} E(X_1) = p$$

as $n \rightarrow \infty$. So,

$$(X_1 + \dots + X_n) \stackrel{\mathcal{D}}{\approx} np \tag{1.13.1}$$

when n is large. Hence, if $n = 10000$ and $p = 0.01$, the approximation number of defectives should be about 100.

The LLN guarantees that even though the "sample average" $n^{-1}(X_1 + \dots + X_n)$ is a rv, it "settles down" to something deterministic and predictable when n is large, namely $E(X_1)$. Hence, even though the individual X_i 's are unpredictable, their average (or mean) is predictable. The fact that the average $n^{-1}(X_1 + \dots + X_n)$ settles down to the expectation $E(X_1)$ is a principal reason for why the expectation of a rv is the most widely used "measure of central tendency" (as opposed, for example, to the median of the distribution).

14 Central Limit Theorem

The second key limit result in probability is the *central limit theorem* (CLT). (It is so important that it is the "central" theorem of probability!)

Note that the LLN approximation (1.13.1) is rather crude:

$$P(X_1 + \dots + X_n \leq x) \approx \begin{cases} 0, & x < np \\ 1, & x \geq np \end{cases}$$

Typically, we'd prefer an approximation that tells us how close $P(X_1 + \dots + X_n \leq x)$ is to 0 when $x < np$ and how close to 1 when $x \geq np$. The CLT provides exactly this additional information.

Theorem 1.14.1 Suppose that the X_i 's are iid rv's with common (finite) variance σ^2 . Then, if $S_n = X_1 + \cdots + X_n$

$$\frac{S_n - nE(X_1)}{\sqrt{n}} \Rightarrow \sigma N(0, 1) \quad (1.14.1)$$

as $n \rightarrow \infty$.

The CLT (1.14.1) supports the use of the approximation

$$S_n \stackrel{\mathcal{D}}{\approx} nE(X_1) + \sqrt{n}\sigma N(0, 1) \quad (1.14.2)$$

when n is large. The approximation (1.14.2) is valuable in many different problem settings. We now illustrate its use with an example.

Example 1.3.3(continued) Let S_{10000} be the number of defectives found in the first 10000 items tested, when $p = 0.01$. The approximation (1.14.2) yields

$$\begin{aligned} P(S_{10000} > x) &\approx P(nE(X_1) + \sqrt{n}\sigma N(0, 1) > x) \\ &= P(N(0, 1) > \frac{x - 100}{100\sigma}), \end{aligned}$$

where $\sigma^2 = \text{var}(X_1) = p(1 - p) = (0.01)(0.99) \approx 0.01$. So,

$$P(S_{10000} > x) \approx P(N(0, 1) > \frac{x}{10} - 10).$$

In particular, the probability of more than 120 defectives is approximately $P(N(0, 1) > 2)$. This later probability can be found in a table of "normal probabilities".

An outline of the proof of the CLT is provided in the next section.

15 Moment Generating Functions*

A key idea of in applied mathematics is that of the Laplace transform. The Laplace transform also is a useful tool in probability. In the probability context, the Laplace transform is usually called the *moment generating function* (of the rv).

Definition 1.15.1 The moment generating function of a rv X is the function $\varphi_X(\theta)$ defined by

$$\varphi_X(\theta) = E(\exp(\theta x)).$$

This function can be computed in "closed form" for many of the distributions encountered most frequently in practice:

- Bernoulli rv: $\varphi_X(\theta) = (1 - p) + pe^\theta$
- Binomial(n, p) rv: $\varphi_X(\theta) = ((1 - p) + pe^\theta)^n$
- Geometric(p) rv: $\varphi_X(\theta) = p/(1 - (1 - p)e^\theta)$
- Poisson(λ) rv: $\varphi_X(\theta) = \exp(\lambda(e^\theta - 1))$

- Uniform(a, b) rv: $\varphi_X(\theta) = (e^{\theta b} - e^{\theta a})/\theta(b - a)$
- Exponential(λ) rv: $\varphi_X(\theta) = \lambda(\lambda - \theta)^{-1}$
- Gamma(λ, α) rv: $\varphi_X(\theta) = \left(\frac{\lambda}{\lambda - \theta}\right)^\alpha$
- Normal(μ, σ^2) rv: $\varphi_X(\theta) = \exp(\theta\mu + \frac{\sigma^2\theta^2}{2})$

The moment generating function (mgf) of a rv X gets its name from the fact that the moments (i.e. $E(X^k)$ for $k = 1, 2, \dots$) of the rv X can easily be computed from knowledge of $\varphi_X(\theta)$. To see this, note that if X is continuous, then

$$\begin{aligned} \frac{d^k}{d\theta^k}\varphi_X(\theta) &= \frac{d^k}{d\theta^k}E(\exp(\theta X)) \\ &= \frac{d^k}{d\theta^k}\int_{-\infty}^{\infty} e^{\theta x} f(x) dx \\ &= \int_{-\infty}^{\infty} \frac{d^k}{d\theta^k} e^{\theta x} f(x) dx \\ &= \int_{-\infty}^{\infty} x^k e^{\theta x} f(x) dx \\ &= E(X^k \exp(\theta X)). \end{aligned}$$

In particular,

$$\frac{d^k}{d\theta^k}\varphi_X(0) = E(X^k).$$

Example 1.15.1 Suppose that X is exponentially distributed with parameter λ . Note that

$$\varphi_X(\theta) = \lambda/(\lambda - \theta)^{-1} = 1/(1 - \frac{\theta}{\lambda})^{-1} = \sum_{k=0}^{\infty} \frac{1}{\lambda^k} \theta^k \quad (1.15.1)$$

On the other hand, $\varphi_X(\theta)$ has the power series representation,

$$\varphi_X(\theta) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{d^k}{d\theta^k}\varphi_X(0)\theta^k \quad (1.15.2)$$

Equating coefficients in (1.15.1) and (1.15.2), we find that

$$\frac{d^k}{d\theta^k}\varphi_X(0) = \frac{k!}{\lambda^k},$$

so that

$$E(X^k) = \frac{k!}{\lambda^k}.$$

Note that we were able to compute all the moments of an exponential rv without having to repeatedly compute integrals.

Another key property of mgf's is the fact that uniquely characterizes the distribution of the rv. In particular, if X and Y are such that $\varphi_X(\theta) = \varphi_Y(\theta)$ for all values of θ , then

$$P(X \leq x) = P(Y \leq x)$$

for all x .

This property turns out to be very useful when combined with the following proposition.

*More mathematically advanced material. 19

Proposition 1.15.1 Let the X_i 's be independent rv's, and put $S_n = X_1 + \cdots + X_n$. Then,

$$\varphi_{S_n}(\theta) = \prod_{i=1}^n \varphi_{X_i}(\theta).$$

Proof: Note that

$$\begin{aligned} \varphi_{S_n}(\theta) &= \mathbb{E}(\exp(\theta(X_1 + \cdots + X_n))) \\ &= \mathbb{E}\left(\prod_{i=1}^n \exp(\theta X_i)\right) \\ &= \prod_{i=1}^n \mathbb{E}(\exp(\theta X_i)) \quad (\text{due to independence}) \\ &= \prod_{i=1}^n \varphi_{X_i}(\theta). \end{aligned}$$

In other words, the mgf of a sum of independence rv's is trivial to compute in terms of the mgf's of the summands. So, one way to compute the exact distribution of a sum of n independent rv's X_1, \dots, X_n is:

- i) Compute $\varphi_{X_i}(\theta)$ for $1 \leq i \leq n$.
- ii) Compute

$$\varphi_{S_n}(\theta) = \prod_{i=1}^n \varphi_{X_i}(\theta).$$

- iii) Find a distribution/rv Y such that

$$\varphi_{S_n}(\theta) = \varphi_Y(\theta)$$

for all θ .

Then,

$$P(S_n \leq x) = P(Y \leq x).$$

Example 1.15.2 Suppose the X_i 's are iid Bernoulli rv's with parameter p . Then,

$$\varphi_{S_n}(\theta) = (1 - p + pe^\theta)^n.$$

But $(1 - p + pe^\theta)^n$ is the mgf of a Binomial rv with parameter n and p . So,

$$P(S_n = k) = P(\text{Bernoulli}(n, p) = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

We conclude this discussion of mgf's by showing how the CLT can be proved by appealing to mgf's. We need the following result.

Theorem 1.15.1 Let $(X_n : 1 \leq n \leq \infty)$ be a sequence of rv's with mgf's $(\varphi_{X_n}(\theta) : 1 \leq n \leq \infty)$. If, for each θ ,

$$\varphi_{X_n}(\theta) \rightarrow \varphi_{X_\infty}(\theta)$$

as $n \rightarrow \infty$, then

$$X_n \Rightarrow X_\infty$$

as $n \rightarrow \infty$.

Proof of the CLT: Let

$$\varphi_n(\theta) = E[\exp(\theta(S_n - E(X_1))/\sqrt{n})]$$

and note that

$$\varphi_n(\theta) = \left[\exp\left(-\frac{\theta}{\sqrt{n}}E(X_1)\right) \varphi_{X_1}\left(\frac{\theta}{\sqrt{n}}\right) \right]^n.$$

But $\varphi_{X_1}(\theta/\sqrt{n})$ can be expanded in a Taylor series about $\theta = 0$:

$$\varphi_{X_1}(\theta/\sqrt{n}) = 1 + E(X_1) \frac{\theta}{\sqrt{n}} + \frac{E(X_1^2)}{2} \frac{\theta^2}{n} + \dots$$

Also, $\exp(-\frac{\theta}{\sqrt{n}}E(X_1))$ can be similarly expanded in a Taylor series:

$$\exp\left(-\frac{\theta}{\sqrt{n}}E(X_1)\right) = 1 - \frac{\theta}{\sqrt{n}}E(X_1) + \frac{E(X_1^2)}{2} \frac{\theta^2}{n} + \dots$$

Hence,

$$\varphi_n(\theta) = \left(1 + \frac{\theta^2}{2n}\sigma^2 + \dots\right)^n \rightarrow \exp\left(\frac{\theta^2\sigma^2}{2}\right)$$

as $n \rightarrow \infty$, which coincides with the mgf of a $N(0, 1)$ rv. Appealing to Theorem 1.15.1 concludes the arguments. \square

16 Application of the LLN and CLT to Financial Mathematics

Many sophisticated financial models (e.g. options pricing) require, as a building block, a model for the value of some underlying financial asset (e.g. the stock that underlies the option). We describe here how the LLN and CLT lead to a particular and widely accepted model for asset values.

Let V_n be the value of the asset at the start of period n . The ratio $R_i = V_i/V_{i-1}$ is called the *return* on the asset of period i . We can then write

$$V_n = V_0 R_1 R_2 \cdots R_n. \tag{1.16.1}$$

A typical (and frequently reasonable) assumption is to presume that the R_i 's are iid random variables. We wish to use the LLN and CLT to develop an approximation to the distribution of V_n when n is large.

Since the LLN and CLT describe the behavior of sums of rv's, we take logarithms in (1.16.1) (thereby concerning the product of rv's to a sum of rv's):

$$\log V_n = \log V_0 + \sum_{i=1}^n \log R_i. \quad (1.16.2)$$

Because the R_i 's are iid, it follows that the $\log R_i$'s are also iid. So, the LLN guarantees that

$$\frac{1}{n} \sum_{i=1}^n \log R_i \xrightarrow{p} \mathbb{E}(\log R_1)$$

as $n \rightarrow \infty$, leading to the approximation

$$\sum_{i=1}^n \log R_i \approx n\mathbb{E}(\log R_1)$$

Hence,

$$\log V_n \approx \log V_0 + n\mathbb{E}(\log R_1)$$

and so,

$$V_n \approx V_0 \exp(n\mathbb{E}(\log R_1)). \quad (1.16.3)$$

The approximation (1.16.3) approximates the random variable V_n by the deterministic quantity $V_0 \exp(n\mathbb{E}(\log R_1))$. While this approximation describes the order of the magnitude of V_n , it fails to characterize the stochastic variability that is present in V_n .

To obtain an approximation that describes the variability, we apply the CLT. According to (1.14.2), we can write

$$\sum_{i=1}^n \log R_i \stackrel{\mathcal{D}}{\approx} n\mathbb{E}(\log R_1) + \sigma\sqrt{n}N(0, 1),$$

where $\sigma^2 = \text{var}(\log R_1)$. It follows that

$$\log V_n \stackrel{\mathcal{D}}{\approx} \log V_0 + n\mathbb{E}(\log R_1) + \sigma\sqrt{n}N(0, 1)$$

when n is large. Consequently, this leads us to the approximation

$$V_n \stackrel{\mathcal{D}}{\approx} V_0 \exp(n\mathbb{E}(\log R_1) + \sigma\sqrt{n}N(0, 1)). \quad (1.16.4)$$

Definition 1.16.1 A nonnegative rv Z is said to be a *log-normal* rv if $\log Z$ is normally distributed.

Note that (1.16.4) asserts that, when n is large, the value V_n is approximately log-normally distributed. In particular, $\log V_n$ is approximately normally distributed with mean $\log V_0 + n\mathbb{E}(\log R_1)$ and standard deviation $\sigma\sqrt{n}$. The above argument justifies the use of log-normal distributions in modeling asset values. In fact, the log-normal distribution is fundamentally linked to the most famous result in financial mathematics, namely, the so-called Black-Scholes formula for valuing so-called European options. This latter result has been recognized with a Nobel Prize in economics.

To apply the approximation (1.16.4), one first computes $\mathbb{E}(\log R_1)$ and σ^2 with these two quantities in hand, (1.16.4) asserts that

$$\begin{aligned} P(V_n > x) &\approx P(V_0 \exp(n\mu + \sigma\sqrt{n}N(0, 1)) > x) \\ &= P(N(0, 1) > \frac{\log x - \log V_0 - n\mathbb{E}(\log R_1)}{\sigma\sqrt{n}}). \end{aligned}$$

One can now use probabilities derived from a “normal table” to approximate $P(V_n > x)$.