Linear Difference Equations of Order One

Given sequences $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$, find a sequence $\{x_n\}_{n=0}^{\infty}$ such that

$$x_n = a_n x_{n-1} + b_n, \quad n = 1, 2, \ldots; \qquad x_0 = c.$$

For the homogeneous case $x_n = a_n x_{n-1}$, the solution is

$$x_n = c\pi_n,$$

where $\pi_0 = 1$, $\pi_n = a_1 a_2 \cdots a_n$. The general (nonhomogeneous) solution is

$$x_n = \pi_n \left(c + \frac{b_1}{\pi_1} + \frac{b_2}{\pi_2} + \cdots + \frac{b_n}{\pi_n} \right).$$

Linear Constant Coefficient Homogeneous Difference Equations of Order k

Given constants $a_1, \ldots, a_k, c_0, \ldots, c_{k-1}$, find a sequence $\{x_n\}_{n=0}^{\infty}$ such that

$$x_n = a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_k x_{n-k}, \quad n \geq k; \qquad x_i = c_i,\ i = 0, 1, \ldots, k-1.$$

For a sequence $X = \{x_n\}_{n=0}^{\infty}$, define the sequence $\mathcal{L}(X) = \{y_n\}_{n=k}^{\infty}$ by

$$y_n = x_n - a_1 x_{n-1} - a_2 x_{n-2} - \cdots - a_k x_{n-k}$$

and the associated characteristic polynomial

$$P(\tau) = \tau^k - a_1 \tau^{k-1} - a_2 \tau^{k-2} - \cdots - a_{k-1}\tau - a_k.$$

Then

(1) $\mathcal{L}(W) = \mathcal{L}(Z) = 0 \Longrightarrow \mathcal{L}(\alpha W + \beta Z) = 0$ for any constants α and β,

(2) for any root t of $P(\tau) = 0$, the sequence X defined by $x_n = t^n$ solves $\mathcal{L}(X) = 0$,

(3) if t is a root of multiplicity $r > 1$ of $P(\tau) = 0$, then $x_n = n(n-1)\cdots(n-s+1)t^{n-s}$ solves $\mathcal{L}(X) = 0$ for $0 \leq s \leq r-1$.

Each simple root t of $P(\tau) = 0$ defines a sequence by $x_n = t^n$, and each multiple root t defines multiple sequences according to (3). There are thus exactly k distinct (and *independent*) sequences $X^{(i)}$, $1 \leq i \leq k$, each solving $\mathcal{L}(X) = 0$. By (1), $\mathcal{L}\left(\sum_{i=1}^{k} \alpha_i X^{(i)}\right) = 0$, and solving the linear system of equations

$$\sum_{i=1}^{k} \alpha_i X_0^{(i)} = c_0,$$

$$\vdots$$

$$\sum_{i=1}^{k} \alpha_i X_{k-1}^{(i)} = c_{k-1},$$

for the coefficients α_i gives the unique solution

$$\sum_{i=1}^{k} \alpha_i X^{(i)}$$

to $\mathcal{L}(X) = 0$ satisfying the initial conditions $x_i = c_i$, $i = 0, 1, \ldots, k-1$.

Example. The Fibonacci sequence is defined by

$$x_n = x_{n-1} + x_{n-2}, \quad x_0 = x_1 = 1.$$

The linear operator $\mathcal{L}(X)_n = x_n - x_{n-1} - x_{n-2}$ and the roots of the characteristic polynomial $P(\tau) = \tau^2 - \tau - 1$ are $\frac{1 \pm \sqrt{5}}{2}$, and thus two solutions of $\mathcal{L}(X) = 0$ are

$$X_n^{(1)} = \left(\frac{1+\sqrt{5}}{2}\right)^n, \qquad X_n^{(2)} = \left(\frac{1-\sqrt{5}}{2}\right)^n.$$

The linear system of equations to satisfy the initial conditions is

$$\begin{aligned} \alpha_1 X_0^{(1)} + \alpha_2 X_0^{(2)} &= \alpha_1 + \alpha_2 = 1, \\ \alpha_1 X_1^{(1)} + \alpha_2 X_1^{(2)} &= \alpha_1 \left(\frac{1+\sqrt{5}}{2}\right) + \alpha_2 \left(\frac{1-\sqrt{5}}{2}\right) = 1, \end{aligned}$$

whose solution is $\alpha_1 = \frac{1+\sqrt{5}}{2\sqrt{5}}$, $\alpha_2 = -\frac{1-\sqrt{5}}{2\sqrt{5}}$. The Fibonacci numbers are therefore given by

$$x_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1} \right).$$

Special Numbers

Euler's constant $\gamma = \lim_{n\to\infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \ln n\right) = 0.57721\,56649\,01532\,86060\ldots$

Euler's gamma function $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}\,dt$, where $\Re z > 0$. $\Gamma(1) = 1$, $\Gamma(n+1) = n!$ for nonnegative integers n, and in general $\Gamma(z+1) = z\Gamma(z)$.

${}_nP_r = \dfrac{n!}{(n-r)!}$ is the number of permutations of n objects taken r at a time.

The *binomial coefficient* $\dbinom{n}{r} = \dfrac{n!}{r!(n-r)!}$, defined for integers $n \geq r \geq 0$, is the number of combinations of n objects taken r at a time. More generally,

$$\binom{n}{r} = \begin{cases} \dfrac{n(n-1)\cdots(n-r+1)}{r!}, & \text{integer } r \geq 0,\\ 0, & \text{integer } r < 0, \end{cases}$$

is defined for real or complex n and integer r, and satisfies the Pascal triangle recurrence

$$\binom{n+1}{r+1} = \binom{n}{r} + \binom{n}{r+1}.$$

By convention, $\binom{n}{0} = 1$.

The *Stirling number of the second kind* $\left\{ {n \atop r} \right\}$ is the number of partitions of a set of n objects into r nonempty subsets. By convention, $\left\{ {n \atop n} \right\} = 1$ for integer $n \geq 0$, and $\left\{ {n \atop r} \right\} = 0$ for integers $n > 0$, $r \leq 0$. The recurrence relation is

$$\left\{ {n+1 \atop r} \right\} = r\left\{ {n \atop r} \right\} + \left\{ {n \atop r-1} \right\}.$$

The *Stirling number of the first kind* $\left[{n \atop r} \right]$ is the number of wreath arrangements of a set of n objects into r nonempty wreaths (or, equivalently, the number of permutations of a set of n objects that can be written as the product of r cycles). For integers $n \geq 0$ and r, some identities are

$$\left[{n \atop n} \right] = \left\{ {n \atop n} \right\} = 1, \qquad \left[{n \atop n-1} \right] = \left\{ {n \atop n-1} \right\} = \binom{n}{2}, \qquad \left[{n+1 \atop 1} \right] = n!, \qquad \sum_{r=0}^{n} \left[{n \atop r} \right] = n!,$$
$$\left[{n+1 \atop r} \right] = n\left[{n \atop r} \right] + \left[{n \atop r-1} \right].$$

Difference Calculus

Difference calculus is the discrete analog of Newton's continuous calculus. For a function $f(x)$, some basic difference operators are

$\Delta f(x) = f(x+1) - f(x)$ (forward difference),

$\nabla f(x) = f(x) - f(x-1)$ (backward difference),

$\delta f(x) = f(x+1/2) - f(x-1/2)$ (central difference),

$\mu f(x) = \big(f(x+1/2) + f(x-1/2)\big)/2$ (averaging operator).

$$x^{\overline{k}} = x(x+1)\cdots(x+k-1), \qquad x^{\underline{k}} = x(x-1)\cdots(x-k+1), \qquad \Delta x^{\underline{k}} = kx^{\underline{k-1}}.$$

$x^{\overline{k}}$ is read as "x^k ascending" and $x^{\underline{k}}$ is read as "x^k descending". Using the fact that $\left\{ {n \atop k} \right\} = \left[{n \atop k} \right] = 0$ for integers $n \geq 0$ and $k < 0$, relationships between x^n, $x^{\underline{n}}$, and $x^{\overline{n}}$ for integer $n \geq 0$ are

$$x^n = \sum_k \left\{ {n \atop k} \right\} x^{\underline{k}} = \sum_k \left\{ {n \atop k} \right\} (-1)^{n-k} x^{\overline{k}},$$

$$x^{\overline{n}} = \sum_k \left[{n \atop k} \right] x^k, \qquad x^{\underline{n}} = \sum_k \left[{n \atop k} \right] (-1)^{n-k} x^k, \qquad x^{\underline{n}} = (-1)^n (-x)^{\overline{n}}.$$

Fundamental Theorem of Difference Calculus. Let $a < b$ be integers, and f, F be functions such that $f(x) = \Delta F(x)$. Then

$$\sum_{a \leq k < b} f(k) = F(b) - F(a)$$

.

Example. Find a formula for $\sum_{k=0}^{n} k^3$. Take

$$f(x) = x^3 = x^{\underline{3}} + 3x^{\underline{2}} + x^{\underline{1}} \text{ and } F(x) = \frac{x^{\underline{4}}}{4} + x^{\underline{3}} + \frac{x^{\underline{2}}}{2}$$

for which $\Delta F(x) = f(x)$. Then by the fundamental theorem,

$$\begin{aligned} \sum_{k=0}^{n} k^3 = \sum_{0 \leq k < n+1} f(k) &= F(n+1) - F(0) \\ &= \frac{(n+1)n(n-1)(n-2)}{4} + (n+1)n(n-1) + \frac{(n+1)n}{2} \\ &= \frac{n^2(n+1)^2}{4}. \end{aligned}$$

Probability and Statistics

A *random variable* is a function $X : A \to B$ defined on a set A of outcomes. e.g., $A = \{\text{heads}, \text{tails}\}$, $B = \{0, 1\}$, $X(\text{heads}) = 1$, $X(\text{tails}) = 0$. Subsets of A are called *events*. Associated with an event $R \subset A$ is a real number $P(R)$, $0 \le P(R) \le 1$, called the *probability* of R, which measures the likelihood of any outcome in R occurring. A probability measure P satisfies

(1) $P(\emptyset) = 0$,

(2) $P(A) = 1$,

(3) $P(R \cup S) = P(R) + P(S)$ for any two disjoint ($R \cap S = \emptyset$) subsets R, S of A.

Example. $A = \{\text{5-card poker hands}\}$, $R = \{\text{hands containing a king}\}$, $S = \{\text{hands with no face cards}\}$. Then $R \cap S = \emptyset$, and $P(R \cup S) = P(R) + P(S)$.

The events R and S are *independent* if $P(R \cap S) = P(R)P(S)$.

For a *discrete* random variable X (the set B is a discrete set, e.g., $B = \{0, 1, 2, 3, \ldots\}$), the probability that $X = x$ is

$$P(X = x) = P\big(\{\theta \mid X(\theta) = x\}\big) = p_X(x).$$

The function p_X is called the *probability mass function* (pmf) of X.

Example. Let $P(\{\text{head}\}) = p$, $A = \{\text{sequences of } n \text{ coin flip outcomes}\}$, $B = \{0, 1, \ldots, n\}$, $\theta \in A$, $X(\theta) =$ number of heads in θ. Then

$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r};$$

X is said to have a *binomial distribution*.

$X : A \to B$ is a *continuous* random variable if A and B are not discrete sets, e.g., $A = B = \mathbf{R} = \{\text{real numbers}\}$. In this case probabilities are given by integrals:

$$P(a \le X \le b) = P\big(\{\theta \mid a \le X(\theta) \le b\}\big) = \int_a^b f(s)\, ds,$$

where $f(s) \ge 0$ is called the *probability density function* (pdf) for X. The function

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(s)\, ds$$

is called the *cumulative distribution function* (cdf) for X; note that $F'(x) = f(x)$.

Some standard continuous probability distributions with their notation and density function are listed below.

Distribution	Notation	pdf $f(x)$	Parameters
uniform	$U(a, b)$	$\dfrac{1}{b-a}$	$a \le x \le b$
normal	$N(\mu, \sigma^2)$	$\dfrac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$	$-\infty < x < \infty,\ \sigma > 0$

gamma	$GAM(\lambda, \alpha)$	$\dfrac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}$	$x > 0,\ \alpha, \lambda > 0$
r-stage Erlang	$GAM(\lambda, r)$		$r = 1, 2, 3, \ldots$
exponential	$EXP(\lambda)$	$\lambda e^{-\lambda x}$	$x > 0,\ \lambda > 0$
hypoexponential	$HYPO(\lambda_1, \ldots, \lambda_n)$	$\sum_{i=1}^{n} a_i \lambda_i e^{-\lambda_i x}$	$x > 0,\ \lambda_i > 0\ \forall i$
		$a_i = \prod_{\substack{j=1 \\ j \neq i}}^{n} \dfrac{\lambda_j}{\lambda_j - \lambda_i}$	$\lambda_i \neq \lambda_j$ for $i \neq j$
beta	$BETA(\alpha, \beta)$	$\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$0 \leq x \leq 1,\ \alpha, \beta > 0$

Define a binary random variable $Y_i(\text{success}) = 1$, $Y_i(\text{failure}) = 0$ for trial i. If the Y_i are independent and $P(Y_i = 1) = p\ \forall i$, then the Y_i/trials are called *Bernoulli* random variables/trials. Each trial may have c different outcomes, with probabilities $p_1, \ldots, p_c$. Some common discrete probability distributions follow.

Distribution	pmf $p_X(r)$	Parameters
binomial	$\binom{n}{r} p^r (1-p)^{n-r}$	r successes in n trials
multinomial	$\dfrac{n!}{r_1! r_2! \cdots r_c!} \prod_{i=1}^{c} p_i^{r_i}$	$r = (r_1, r_2, \ldots, r_c)$ outcomes in $n = \sum_{i=1}^{c} r_i$ trials
geometric	$(1-p)^{r-1} p$	r trials up to and including first success
Poisson	$\dfrac{e^{-\mu} \mu^r}{r!}$	$\mu > 0$, $r = 0, 1, 2, \ldots$

The *expected value* (or *mean*) $E[X]$ of a discrete random variable X with pmf $p_X(x)$ is defined by

$$E[X] = \sum_{x \in B} x\, p_X(x),$$

and for a continuous random variable X with pdf $f(s)$ by

$$E[X] = \int_{-\infty}^{\infty} s f(s)\, ds.$$

In general, the expected value of any function $g(X)$ of a random variable X is

$$E[g(X)] = \sum_{x \in B} g(x)\, p_X(x) \text{ or } E[g(X)] = \int_{-\infty}^{\infty} g(s) f(s)\, ds.$$

The *variance* of X is $Var[X] = E\big[\big(X - E[X]\big)^2\big]$. The mean and variance are often denoted by μ and σ^2, respectively. σ is called the *standard deviation*.

Theorem. For any two random variables X and Y, $E[X + Y] = E[X] + E[Y]$.

Theorem. If X and Y are independent random variables, then

$$E[XY] = E[X]E[Y] \quad \text{and} \quad Var[X+Y] = Var[X] + Var[Y].$$

Example. For a Bernoulli variable Y_i, $E[Y_i] = 1 \cdot p + 0 \cdot (1-p) = p$, $Var[Y_i] = E[(Y_i - p)^2] = E[Y_i^2] - (E[Y_i])^2 = p - p^2 = p(1-p)$.

Example. For an n-trial binomial variable $X = Y_1 + \cdots + Y_n$, $E[X] = E\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} E[Y_i] = np$ and $Var[X] = Var\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} Var[Y_i] = np(1-p)$.

Example. X with a geometric distribution has $E[X] = 1/p$ and $Var[X] = (1-p)/p^2$.

Example. For a Poisson variable X with parameter μ, $E[X] = Var[X] = \mu$.

Example. A uniform distribution X over $a \le x \le b$ has $E[X] = (a+b)/2$ and $Var[X] = (b-a)^2/12$.

Example. An $N(\mu, \sigma^2)$ normal random variable has mean μ and variance σ^2.

Example. A $GAM(\lambda, \alpha)$ random variable has mean α/λ and variance α/λ^2.

Markov Inequality. Let X be a nonnegative random variable with finite mean $E[X] = \mu$. Then for any $t > 0$,

$$P(X \ge t) \le \frac{\mu}{t}.$$

Chebyshev Inequality. Let X be a random variable with finite mean μ and variance σ^2. Then for any $t > 0$,

$$P\big(|X - \mu| \ge t\big) \le \frac{\sigma^2}{t^2}.$$

Consider a population described by a distribution (with mean μ and variance σ^2), where a value $X(\theta) = x$ of the random variable X corresponds to sampling one individual from the population. A sample $x_1, \ldots, x_n$ from the population can be viewed as values of n independent identically distributed random variables $X_1, \ldots, X_n$. A *statistic* is a number derived from a sample or population, and the fundamental question is how sample statistics relate to population statistics.

The *sample mean* $\bar{x}$ and *sample variance* $s^2 = \bar{\sigma}^2$ are

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}, \qquad s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

These are called *unbiased* estimators, since $E[\bar{x}] = \mu$ and $E[s^2] = \sigma^2$. Observe that if s^2 were defined with $1/n$ instead of $1/(n-1)$, then $E[s^2] \ne \sigma^2$. Applying Chebyshev's inequality gives

$$P\big(|\bar{x} - \mu| \ge t\big) \le \frac{\sigma^2}{nt^2}.$$

Let S be an *event space* (set of outcomes), let $A, B \subset S$ be events, and let the events B_1, B_2, …, $B_n \subset S$ partition S, i.e., $S = \cup_{i=1}^{n} B_i$ and $B_i \cap B_j = \emptyset$ for $i \neq j$. The *conditional probability* of event B, given that event A has already occurred, is defined by

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)}, \qquad P(A) > 0.$$

For $P(A) > 0$, *Bayes Theorem* states that for each $k = 1, \ldots, n$,

$$P\big(B_k \mid A\big) = \frac{P\big(A \mid B_k\big)\, P\big(B_k\big)}{\displaystyle\sum_{i=1}^{n} P\big(A \mid B_i\big)\, P\big(B_i\big)}.$$