A

1-randomness *see* Martin-Löf randomness

Abelian sandpile (alias: **Bak–Tang–Wiesenfeld model; sandpile automaton; lattice avalanche model; Dhar's Abelian sandpile**) A classic interacting particle system, where the particles are envisaged as grains of sand that form piles at the points of a subset D of the d-dimensional integer lattice. Each vertex can support grains up to some finite limit; at integer times a new grain is added to a randomly chosen point $v \in D$. If this addition causes the pile to exceed its limit, it will topple onto its neighbours according to some rule of redistribution. This **toppling** may induce further topplings, and a complete set of topplings is called an **avalanche**. The process is called Abelian when the final stable result of an avalanche is independent of the order of execution of the topplings. No grains are created in avalanches, but sand may be lost at a boundary. Other sandpile models exist, e.g. **Zhang's sandpile** in which a random quantity of sand, continuously distributed on an interval $[a, b]$, is added to a randomly chosen vertex. If its capacity is exceeded, the pile is equally divided among all the vertex's neighbours. Related models include the Bak–Sneppen model, the chip-firing game, and the Dirichlet game. [G. Pruessner, *Self-Organized Criticality*. Cambridge University Press, 2012]

absolute continuity A strengthening of the idea of continuity that arises in the context of real-valued functions and measures. One definition asserts that a real function $F(x), x \in \mathbb{R}$ is absolutely continuous if it can be written as the Lebesgue integral of a function $f(x)$; $F(x) = \int_a^x f(u)\mathrm{d}u$. Thus the distribution of a random variable having a density on $\mathbb{R}$ (or $\mathbb{R}^d$), is absolutely continuous. Equivalently, $F(x)$ is absolutely continuous on $[a, b]$ if, for any $\varepsilon > 0$, there is $\delta > 0$ such that $\sum_k |F(b_k) - F(a_k)| < \varepsilon$, for any collection of disjoint intervals $(a_k, b_k) \in [a, b]$, such that $\sum_k (b_k - a_k) < \delta$.

A classic example of a continuous, yet not absolutely continuous, distribution function is the Cantor distribution. Similarly, but more generally, a measure (or distribution) $\mu(\cdot)$ is said to be absolutely continuous with respect to a measure (or distribution) $\nu(\cdot)$ if $\mu(A) = \int_A f(x)\,\nu(\mathrm{d}x)$ for some $f(\cdot)$. When μ and f are probability distributions, f is called a **density** of μ with respect to ν. More generally, f is called a Radon–Nikodym derivative of μ with respect to ν.

The Radon–Nikodym theorem (sometimes called the Lebesgue–Nikodym thm) shows that μ is absolutely continuous wrt ν iff $\mu(A) = 0$ whenever $\nu(A) = 0$. In this case one says that μ is dominated by the **dominating measure** ν, and writes $\mu \ll \nu$. If also $\nu \ll \mu$, then ν and μ are said to be **equivalent measures**.

The concepts coincide for a distribution function $F(x)$ corresponding to a probability measure $\mathbb{P}(\cdot)$ on $\mathbb{R}$ such that $\mathbb{P}((-\infty, x]) = F(x)$; $\mathbb{P}(\cdot)$ is absolutely continuous wrt the Lebesgue measure iff $F(\cdot)$ is an absolutely continuous function. [D. Pollard, *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, 2002]

absolute deviation The absolute value $|X - q|$ of the difference $X - q$ between a random variable X and some given q, commonly called the mean, the median, or the 'true value' of some quantity, of which X is a sample or realization. It is sometimes called the absolute error; and $|X - q|/q$ may be called the **relative deviation** (or error). [M. Bean, *Probability: The Science of Uncertainty*. American Mathematical Society, 2001]

absolute moment The rth absolute moment of the r.v. X is $\mathbb{E}(|X|^r) = \mathbb{E}(|X^r|)$. If r is not an integer then this is the absolute fractional moment; note that for z complex and r real, $z^r = |z|^r e^{ir\theta}$, where θ is the argument of z. The first absolute moment may be called the **mean (absolute) deviation**. [K. L. Chung, *Course in Probability Theory*, 2nd edn revised. Academic Press, 2001]

absolute monotonicity The probability generating function $G(x)$ of a non-negative integer-valued r.v. has a power series expansion with positive coefficients for $0 \le x < 1$; any function with this property is said to be absolutely monotone. From a theorem of S. Bernstein, this holds iff

$$\sum_{r=0}^{k} \binom{k}{r} (-1)^{k-r} G(rh) \ge 0, \quad 0 \le k \le n-1, h = n^{-1}, \quad n \ge 1.$$

[W. Feller, *Introduction to Probability Theory and its Applications*, Vol. 2, 2nd edn. Wiley, 1971]

Cambridge University Press
978-1-107-07516-0 - The Cambridge Dictionary of Probability and Its Applications
David Stirzaker
Excerpt
More information

absolute probability This is used to denote the unconditional probability of an event, where the lack of conditioning is to be stressed. For example, for a Markov chain X_n with initial distribution a_r and n-step transition probabilities $p_{ij}(n)$, the absolute probability of the event $X_n = j$ is $\sum_r a_r p_{rj}(n)$; in contrast to the conditional $p_{ij}(n)$. Similarly, the term **absolute probability fn** is used in an axiomatization of probability defined on the statements of a propositional language; where the use of 'absolute' stresses the distinction from **relative probability fns**. These are functions of two propositions, one of which is a conditioning statement seen as evidence or support for the other. [W. Feller, *Introduction to Probability Theory and its Applications*, Vol. 1, 3rd edn. Wiley, 1968]

absolute regularity *see* mixing

absorbing A term arising in the context of random processes whose evolution can be seen as the movement of a particle in some space. A point is absorbing if the particle remains there for all time after its first arrival there. Such a point (or state of the process) may be called a **sink** or **trap**.

A collection of such points may be called an **absorbing barrier**, or absorbing boundary. They are important not only as models of reality, but also because all first-passage problems can be formulated in terms of the original process together with additional (artificial) absorbing barriers.

A related concept is that of the **absorbing set** S, which the process can never leave after entering for the first time, but may continue its evolution within S. [S. Karlin & H. Taylor, *A First Course in Stochastic Processes*, 2nd edn. Academic Press, 1975; *A Second Course in Stochastic Processes*. Academic Press, 1981]

absorption-time distribution *see* phase-type distribution

accelerated life model Arises in reliability theory (and other contexts) when the failure-time distribution has a heavy tail (or a large mean) so that empirical results are delayed and expensive. The core of the technique is the assumption that the parameters of the model depend on some measure of stress (e.g. temperature or loading). Data may be obtained quickly and cheaply under high stress and used, via the model, to predict low-stress lifetimes. [V. Bagdonavičius & M. Nikulin, *Accelerated Life Models*. Chapman & Hall, 2002]

acceptability A finite set $X_1, \ldots, X_n$ of r.vs is called acceptable if for any $t \in \mathbb{R}$, $\mathbb{E}\exp\left[t\sum_1^n X_r\right] \le \prod_1^n \mathbb{E}\exp(tX_r)$; while a sequence $X_1, X_2, \ldots$ is acceptable if any finite subset is acceptable. [R. Giuliano Antonini *et al.*, Convergence of series of dependent φ-subgaussian random variables, *J. Math. Anal. Appl.*, 338(2), 1188–1203, 2008]

acceptance-complement A method for sampling from a density $f_X(x)$, as follows. Choose non-negative functions $f_1(x)$ and $f_2(x)$, and a proper density $f_Y(x)$, such that for $x \in \mathbb{R}$ $f_X(x) = f_1(x) + f_2(x)$, and $f_1(x) \le f_Y(x)$. Then:

1. Generate Y with density $f_Y(y)$, and independent U uniform on $[0, 1]$.
2. If $U \le f_1(Y)/f_Y(Y)$, then set $X = Y$; otherwise generate Z with density $f_2(z)/\int_{\mathbb{R}} f_2(x)\,dx$, and set $X = Z$. The output X has density $f_X(x)$.

[R. Kronmal & A. Peterson, Jr, An acceptance-complement analogue of the mixture-plus-acceptance-rejection method for generating random variables, *ACM Trans. Math. Software*, 10(3), 271–281, 1984]

acceptance-rejection (alias: **rejection method; Von Neumann's method**) A method for sampling from a density $f_X(x)$, as follows. Choose a density $f_Y(y)$ such that for some constant a, and all $x, 0 \le f_X(x) \le af_Y(x)$.

1. Generate independent U and Y, where Y has density $f_Y(y)$, and U is uniform on $(0, 1)$.
2. If $aUf_Y(Y) \le f_X(Y)$, then set $X = Y$. Otherwise reject Y and go to step 1.

[R. Rubinstein & D. Kroese, *Simulation and the Monte Carlo Method*, 2nd edn. Wiley, 2008]

acceptance sampling A class of quality control methods that decide whether to accept the output of some industrial (or other) process as satisfactory (or reject it as unsatisfactory) on the basis of a random sample from each lot. There are two main types of sentencing rule (i.e. the decision function for acceptance or rejection of the lot):

1. A plan (or rule) that depends on the number of defective (or non-conforming) items is said to be acceptance by attributes.
2. A plan that depends on measurement of some variable property of the items is acceptance by variables.

More complicated procedures are typically sequential, rather than simply lot-based.
[E. Schilling & D. Neubauer, *Acceptance Sampling in Quality Control*, 2nd edn. Chapman & Hall, 2009]

accessibility A concept arising in the context of Markov chains. Let $X_n \in S$ be a chain with discrete state space S, with $i, j \in S$. Then j is accessible from i if, for some finite $n, \mathbb{P}(X_n = j \,|\, X_0 = i) > 0$. Equivalently, if T_{ij} is the hitting time of j from i, then j is accessible from i if $\mathbb{P}(T_{ij} < \infty) > 0$.

We write $i \to j$ for this property, and synonyms are: i **leads to** (or communicates with) j; j can be reached (or is **reachable**) from i; j is a **consequent** of i.

If $i \to j$ and $j \to i$, then i and j are said to **communicate** (or inter-communicate); and we write $i \leftrightarrow j$ for this property, which is an equivalence relation. If the state space of a chain comprises exactly one such equivalence class, then it is called **irreducible**.

For a Markov chain in continuous time with standard transition probabilities $p_{ij}(t)$, the **Lévy dichotomy** asserts that either $p_{ij}(t) > 0$ for all $t > 0$, or $p_{ij}(t) = 0$ for all t. If $p_{ij}(t)$ have a stable Q-matrix Q, and $\hat{p}_{ij}(t)$ are the transition probabilities for the associated Feller-minimal process $\hat{X}(t)$, then $i \to j$ for $\hat{X}(t)$ iff $i \to j$ for the jump chain defined by Q. For diffusions complications arise, but it is often convenient to say that a state (or set of states) is accessible from x_0 if its occupation measure is greater than zero when started from x_0.

The term also arises in the general theory of processes, where a stopping time T is called accessible if there is a sequence $T_1, T_2, \ldots$ of predictable times such that $\mathbb{P}(\cup_{r=1}^{\infty}\{T_r = T < \infty\}) = \mathbb{P}(T < \infty)$, in which case the sequence is said to catch (or envelop) T. [O. Kallenberg, *Foundations of Modern Probability*, 2nd edn. Springer, 2002]

access time Arises in the context of Markov chains as a generalization of hitting times. Let a finite irreducible Markov chain be started with distribution p, and consider all rules that stop the chain at a time when it has distribution q. The minimum over all such rules of the expected stopping time of the chain is called the **access time** from p to q, denoted by $A(p, q)$. If p and q are degenerate on states i and j respectively, then the access time is just the hitting time of j from i.

If π is the stationary distribution of the chain, and $\delta(i)$ is the degenerate distribution on the state i, then $\max_i A(\delta(i), \pi)$ is called the **mixing time**. Also the quantity $\sum_i \pi_i A(\delta(i), \pi)$ is called the **reset time**. Finally, the **forget time** is $\min_\tau \max_i A(\delta(i), \tau)$, and the distribution τ that achieves this minimum is the forget distribution. [L. Lovász & P. Winkler, Reversal of Markov chains and the forget time, *Combin. Probab. Comp.*, 7, 189–204, 1998]

activated random walk An interacting particle system on the vertices of a suitable graph. Initially a number of particles are distributed (possibly randomly) on the vertices of the graph; they are all active, which is to say that each performs a symmetric r.w. on the vertices independently of the others. Each particle becomes inactive (falls asleep) at rate λ, and once inactive (sleeping) it stops walking until another (necessarily active) particle is present at the same vertex, whereupon it resumes the random walk. In particular, no particle can fall asleep at a vertex with two or more particles present. If there is a time T after which a vertex V is never again visited by an active particle, then V is said to **fixate** at fixating time T. *See also* frog model. [R. Dickman *et al.*, Activated random walkers: Facts, conjectures and challenges, *J. Stat. Phys.*, 138, 126–142, 2010]

actuarial table *see* life table

adapted process (alias: **non-anticipating process**) It is natural to assume that for a random process evolving in time, its distribution at time t (given its past) should not depend on its future values. The adapted process makes this idea formal, thus: given a filtration $(\mathcal{F}_t)$, a random process $X(t)$ is adapted to $(\mathcal{F}_t)$ if, for all t, $X(t)$ is $\mathcal{F}_t$-measurable. Any process is adapted to its natural or induced filtration $(\mathcal{F}_t) = \sigma\{X(s) : s \leq t\}$.

In continuous time, a right-(or left-)continuous adapted process $X(t)$ has the stronger property of being progressively measurable. In discrete time an adapted process is both progressively measurable and an optional process. [D. Applebaum, *Lévy Processes and Stochastic Calculus*, 2nd edn. Cambridge University Press, 2009]

adaptive rejection A method for sampling from a density $f_X(x)$ that is a computationally efficient modification of the acceptance-rejection method. Whenever a rejection occurs, the bounding function $af_Y(x)$ is modified to reduce the number of future rejections. [W. Gilks *et al.*, Adaptive rejection Metropolis sampling, *Appl. Stat.*, 44, 455–472, 1995]

adaptive stochastic control Broadly, that part of control theory that seeks to optimize some function of the realizations of a random process with uncertain parameters, in which the control policy is adapted in the light of updated estimates of the state variables and the unknown parameters.

addition rule (alias: **additive property**) The axiom of classical probability theories which asserts that the probability of the union of disjoint events is the sum of their probabilities. Formally, if $A \cap B = \emptyset$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

This rule is natural when probability is interpreted as frequency, or proportion, or an expression of symmetry, but its necessity has been questioned for interpretations such as degree of belief, or degree of support.

The special form of the rule, written for arbitrary A and B as $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B)$, is often called the **partition rule** or the **law of total** (or **complete**) **probability**. More strongly, Kolmogorov's additivity axiom requires that for a countable pairwise disjoint collection of events A_i, $\mathbb{P}(\bigcup A_i) = \sum \mathbb{P}(A_i)$, and $\mathbb{P}(\cdot)$ is said to be **countably additive** or **sigma-additive**. *See* inclusion-exclusion bounds. [H. Tijms, *Understanding Probability*, 3rd edn. Cambridge University Press, 2012]

additive functional Arises in the context of a strong Markov process $X(t) = X(t, \omega)$. A continuous additive functional of (or associated with) $X(t, \omega)$ is a random process $A(t, \omega)$ such that $A(0, \omega) = 0$, and for $s, t \geq 0$, $A(t+s, \omega) = A(t, \omega) + A(s, \omega_t^+) = A(t, \omega) + A(s, \omega) \circ \theta_t$; where ω_t^+ is the sample path shifted by t (i.e. seen after t) and θ_t is the shift operator. Furthermore, $A(t, \omega)$ is required to be continuous, non-decreasing, and adapted to the filtration generated by $X(t, \omega)$. A classic example is $A(t) = \int_0^t f(X(u))\mathrm{d}u$, where $f(\cdot)$ is a bounded non-negative function; another example is the local time of a diffusion at x. Some writers do not require the functional to be non-decreasing. A **multiplicative functional** is defined analogously by $M(t+s) = M(s)\,M(t) \circ \theta_s$. [S. Karlin & H. Taylor, *A Second Course in Stochastic Processes*. Academic Press, 1981]

additive process (alias: **independent (or orthogonal) increments process**) If, for any n, and $t_1 < t_2 < \cdots < t_n$, the increments $X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})$ of $X(t)$ are independent, then $X(t)$ is said to be an additive (or strongly additive) process. If the increments are only uncorrelated, then the process is said to be weakly additive, or to have orthogonal increments. The Wiener, Lévy, and Poisson processes are classic examples of additive processes. [K. Itô, *Essentials of Stochastic Processes*, English translation. American Mathematical Society, 2006]

additive property *see* addition rule

additive set functional *see* valuation

Adelson recursion A formula for the computation of the terms of a compound Poisson distribution, recursively. Thus, if $X = \sum_{r=1}^N Y_r$ where N is Poisson (λ), and the Y_r are i.i.d.r.v. with $\mathbb{P}(Y_r = k) = p_k, k \geq 0$, then $\mathbb{P}(X = 0) = \exp\{-\lambda(1 - p_0)\}$, and $x\mathbb{P}(X = x) = \lambda \sum_{k=0}^{x-1} (x-k) p_{x-k}\, \mathbb{P}(X = k), \quad x \geq 1$. The case $p_1 = 1$ is the classic Poisson recursion., and a generalization is called the **Panjer** (or **Adelson–Panjer**) **recursion**. [R. Adelson, Compound Poisson distributions, *Oper. Res. Q.*, 17(1), 73–75, 1966]

adiabatic effect Arises in the context of Monte Carlo simulations; it refers to the possibility that an over-hasty annealing schedule may leave the system (process) in a locally extreme (metastable) state, rather than the desired global extreme (ground state). [B. Gidas, Metropolis-Type Monte Carlo simulation algorithms and simulated annealing, in *Topics in Contemporary Probability and its Applications*, L. Snell (ed.). CRC Press, 1995]

adjoint source In communication theory, that memoryless discrete source which has the same distribution over its symbols as a given ergodic Markov source in equilibrium. [D. Welsh, *Codes and Cryptography*. Oxford University Press, 1988]

Adleman's theorem Addresses randomized algorithms in the context of computational complexity. It asserts that the set of decision problems solvable by a probabilistic Turing machine in polynomial time (with error probability less than $p \in (0, 1/2)$ is included in the set of problems decidable by a Turing machine whose circuits and running time are polynomially bounded as functions of the length of the input, denoted by P/poly. [R. Motwani & P. Raghavan, *Randomized Algorithms*. Cambridge University Press, 1995]

admission control Arises in queueing theory, or related systems, when the arrival process of customers (or items) depends on the number, and possibly types, of those in service. The control may be random, as e.g. in balking and reneging, or it may be imposed with the aim of optimizing some measure of performance. [S. Stidham & R. Weber, A survey of Markov decision models for control of networks of queues, *Queueing Sys.*, 13(1–3), 291–314, 1993]

adversarial input (alias: **worst-case**) A concept arising in game theory that has applications in areas as disparate as cryptography and sequential job scheduling in OR. Broadly, it is the assumption that the random inputs to the system are in some sense worst-case; that is, as if chosen by an adversary.

affine Markov process A term arising in two contexts:

1. Given a Markov chain X_n, if there exist suitable functions f and g such that, for all $n \geq 0, X_{n+1} = g(X_n) + f(X_n)Z_{n+1}$, where the Z_n are i.i.d.r.vs independent of X_0, then the chain is said to be affine.
2. If the transition semi-group of a Markov process in continuous time is such that the logarithm of its characteristic function is an affine transformation of the initial state, then the process is called affine.

[T. Hurd & A. Kuznetsov, Affine Markov chain model of multifirm credit migration, *J. Credit Risk*, 3(1), 3–29, 2007; D. Dawson & Z. Li, Skew convolution semigroups and affine Markov processes, *Ann. Probab.*, 34(3), 1103–1142, 2006]

affine process *see* harness

age (alias: **current life; spent waiting time; backward recurrence time**) In a renewal process at time t, the time elapsed since the most recent prior event is called the age of the process. If there is no prior event the age is taken to be t, for a process started at 0. If the interevent time is not arithmetic, with distn fn $F(x)$, then the limiting distn of the age as $t \to \infty$ is $A(x) = (\mathbb{E}X)^{-1}\int_0^x (1 - F(u))\mathrm{d}u$, provided $\mathbb{E}X = \int_0^\infty (1 - F(u))\mathrm{d}u$ is finite. [G. R. Grimmett & D. R. Stirzaker, *Probability and Random Processes*, 3rd edn. Oxford University Press, 2001]

age-dependent process (alias: **Bellman–Harris process**) Generalizes the ordinary Galton–Watson process by requiring each individual to complete a random lifetime before branching. In the simplest model, lifetimes are independent of each other and the family sizes. More generally, a family-size distribution may depend on the age of the individual giving rise to it. [P. Haccou *et al.*, *Branching Processes*. Cambridge University Press, 2005]

ageing Arises in the analysis of reliability and maintenance of systems (in the most general sense), as well as in demographic and actuarial theories. In the simplest classical model, a system comprises n components (usually satisfying some dependency relations), each of which has a random lifetime. For a component aged t, whose lifetime distribution is $F(x)$, for a lifetime T with density $f(x)$, we define $r(t) = f(t)/\{1-F(t)\}$, which is called either the **failure rate fn** or the **hazard rate fn** (or the **force of mortality** in actuarial contexts).

There is a considerable taxonomy based on the properties of $r(t)$, and suitable functions of $r(t)$. The **cumulative hazard fn** is $H(t) = \int_0^t r(u)\mathrm{d}u = -\log(1 - F(t))$(also, and confusingly, sometimes called the hazard function), and for a random lifetime T, the residual lifetime $R(t)$ at t is the random variable $T - t$, conditional on the event $T > t$. Then these classifications are commonplace:

(A) If $H(t)$ is convex (resp. concave) in t, then T is said to have increasing (resp. decreasing) failure rate, denoted by IFR (resp. DFR).
(B) If $t^{-1}H(t)$ is non-decreasing (resp. non-increasing) then T is said to have increasing (resp. decreasing) failure rate on average, denoted by IFRA (resp. DFRA).
(C) If for all $u, v \geq 0, H(u+v) \geq H(u) + H(v)$(resp. $\leq$), then T is said to be new better (resp. worse) than used, denoted by NBU (resp. NWU).
(D) If $\mathbb{E}T \geq \mathbb{E}R(t)$(resp $\leq$), then T is said to be new better (resp. worse) than used in expectation, denoted by NBUE (resp. NWUE).
(E) If $\mathbb{E}R(t)$ is non-increasing (resp. non-decreasing) in t, then T has decreasing (resp. increasing) mean residual life, denoted by DMRL (resp. IMRL).

It can be shown that IFR $\Rightarrow$ DMRL $\Rightarrow$ NBUE and that IFR $\Rightarrow$ IFRA $\Rightarrow$ NBU $\Rightarrow$ NBUE.

The equivalent expression, $1 - F(t) = \mathrm{e}^{-H(t)} = \exp\{-\int_0^t r(u)\mathrm{d}u\}$, may be called the **exponential formula for reliability** in appropriate contexts. More generally, when F has no density, the hazard rate over $(t, t+a], a > 0$, is

$$r(t,a) = \frac{F(t+a) - F(t)}{1 - F(t)} = \mathbb{P}(t < X \leq t + a \,|\, X > t).$$

More narrowly, if X is integer-valued, then $r(t) = \mathbb{P}(X = t \,|\, X \geq t)$. [M. Finkelstein, *Failure Rate Modelling for Reliability and Risk*. Springer, 2008]

age replacement model (or policy) Any one of a large number of procedures for maintaining the reliability of a system by replacing components according to rules based on their current lifetime, or age. [T. Nakagawa, *Maintenance Theory of Reliability*. Springer, 2005]

age-structured process *see* age-dependent process

aggregation *see* lumpability

aggregation-disaggregation A method sometimes used in the analysis of Markov chains arising in the context of very large networks (of queues, or processors). If the state space can be partitioned into subsets within which transitions are very likely, but transitions into other subsets are very unlikely, then one may proceed by first considering inter-set transitions, and then intra-set transitions, to yield good approximate results. [P. Schweitzer, A survey of aggregation-disaggregation in large Markov chains, in *Numerical Solution of Markov Chains*, W. Stewart (ed.). Marcel Dekker, 1991]

aggregation paradox Refers to the fact that it is possible that a number of distinct and disparate (but commensurate) data-sets may all point towards some common conclusion, but when combined (i.e. pooled) the aggregated data-sets suggest the opposite conclusion. *See* Yule–Simpson paradox. [D. Saari, A chaotic exploration of aggregation paradoxes, *SIAM Rev*., 37(1), 37–52, 1995]

Agrawal–Biswas test A randomized method for testing the primality of a given integer n. It relies on the fact that for any $c > 1$, such that $\gcd(n, c) = 1$, n is prime iff $(x - c)^n = x^n - c$, modulo n. Broadly, the method chooses a random monic polynomial $R(x)$ of degree $\lceil \log n \rceil$ = smallest integer not less than $\log n$. If $(x + 1)^n \neq x^n + 1$, modulo $(R(x), n)$, then n is declared composite; otherwise n may be prime. [L. Rempe-Gillen & R. Waldecker, *Primality Testing for Beginners*. American Mathematical Society, 2014]

agreeing function Arises in a number of contexts, e.g. decision theory and theories of probability, where a family T of objects are subject to some ordering induced by, e.g., preferences or beliefs. The real-valued function $a(\cdot)$ is said to agree with this ordering if, for all $x, y \in T, a(x) \leq a(y)$ iff y is preferred to x (or belief in y is greater than belief in x). *See* utility functions. [J. Halpern, A counterexample to theorems of Cox and Fine, *J. Artif. Intel. Res*., 10, 67–85, 1999]

Ahlswede–Daykin inequality (alias: **four functions theorem**) This has applications in random graphs, and other areas of probabilistic combinatorics. One form is this: Let π be the set of all subsets of $\{1, 2, \ldots, n\}$ (i.e. its power set), and let $a(\cdot), b(\cdot), c(\cdot)$, and $d(\cdot)$ be four non-negative functions on π. For a family ϕ of subsets of π, the sum $\sum_{S\in\phi} a(S)$ is denoted by $A(\phi)$, and likewise for b, c, and d.

If, for any two subsets S and T of π, we have $a(S)b(T) \leq c(S \cup T)\, d(S \cap T)$, then for any two families ϕ and θ of subsets of π, $A(\phi)B(\theta) \leq C(\phi \cup \theta)D(\phi \cap \theta)$, where $\phi \cup \theta = \{A \cup B : A \in \phi,\ B \in \theta\}$, and similarly for $\phi \cap \theta$.

It is an example of a correlation inequality, such as those of Harris, FKG, Holley and Fishburn–Shepp, all of which it implies. [R. Ahlswede & D. Daykin, An inequality for the weights of two families of sets, their unions and intersections, *Probab. Theor. Rel. Fields*, 43(3), 183–185, 1978]

Airy distribution Usually denotes one of two distributions:

1. The **map-Airy** distribution, so called because it arises in the study of random maps (i.e. planar graphs), which has density on $\mathbb{R}$ given by $f(x) = 2[xA(x^2) - A'(x^2)]\exp\left[-\frac{2}{3}x^3\right]$, where $A(x) = \pi^{-1} \int_0^\infty \cos\left(xt + \frac{1}{3}t^3\right)dt$, is the Airy function. It has zero mean, but no higher moments.
2. The **area-Airy** distribution, so called because it is the distribution of the area (suitably scaled) of the positive Brownian excursion, which may be defined on $\mathbb{R}^+$ by its moments m_r given by $m_r = A_r 2\sqrt{\pi}/\Gamma\left(\frac{1}{2}(3r - 1)\right), \quad r \geq 1$, where A_r are the Airy constants $\{A_1 = 1/2, A_2 = 5/4, A_3 = 45/4, \ldots\}$. It arises as the limiting distribution of a number of random structures.

[C. Banderier & G Louchard, *Philippe Flajolet and The Airy Function*, online, www.stat.purdue.edu/~mdw/ChapterIntroductions/banderier-louchard.pdf

Aldous–Broder method A procedure for generating a uniform spanning tree on a finite connected graph G (i.e. one selected uniformly at random from all spanning trees). From any vertex v run a symmetric random walk on G until all vertices have been visited. For any vertex $u \neq v$, let $e(u)$ be the edge used to make the first visit to u. The collection of such edges is the required random tree. [S. Evans *et al.*, Rayleigh processes, real trees, and root growth with regrafting, *Probab. Theor. Rel. Fields*, 134(1), 81–126, 2006]

Aldous's condition Arises in considering the weak convergence of a sequence of random processes, where it supplies a sufficient condition for tightness in terms of behaviour after stopping times; a later weaker condition is similarly denoted. [D. Aldous, Stopping times and tightness

I & II, *Ann. Probab.*, 6(2), 335–340, 1978; 17(2), 586–595, 1989]

Aldous's integrated superBrownian excursion (or ISE) Introduced by D. Aldous as a description for random distribution of masses, this is a random probability measure arising as a continuous rescaled limit of random trees. [D. Aldous, Tree-based models for random distribution of mass, *J. Stat. Phys.*, 73(3–4), 625–641, 1993]

aleatory Random. From the Latin *aleator*, meaning one who throws a die; the Latin word for die being *alea* (which was also the name of a popular Roman board game).

The term 'aleatory probability' may be applied in the context of uncertain physical events lying in the future, such as the decay of subatomic particles or radioactive atoms; it is to be contrasted with epistemic probability, which arises simply from a lack of knowledge (see probability). [I. Hacking, *The Emergence of Probability*. Cambridge University Press, 1975]

Alexandrov's theorem *see* portmanteau theorem

algebraic probability *see* quantum probability

algorithmic complexity *see* Kolmogorov complexity

algorithmic probability (alias: **Solomonoff probability; universal probability**) Commonly denotes a probability distribution assigned to binary strings (finite or infinite) on an a priori basis, i.e. axiomatically not empirically (as the uniform distribution is similarly assigned by the principle of insufficient reason). The concept was introduced by R. Solomonoff for the purposes of inductive inference, and is closely related to Kolmogorov complexity. In the basic case of a finite binary string b the so-called universal (algorithmic) probability $\mathbb{P}(b)$ of b is defined in terms of a universal (Turing) computer with random (prefix-free) input p having length $l(p)$ yielding output $T(p)$.

Then $\mathbb{P}(b) = \sum_{p:T(p)=b} 2^{-l(p)}$, where $\sum_p 2^{-l(p)} = \Omega_c < 1$ is Chaitin's constant. Note that the term algorithmic probability also has its standard interpretation. [M. Hutter, *Universal Artificial Intelligence*. Springer, 2005; M. Neuts, *Algorithmic Probability*. Chapman & Hall, 1995]

algorithmic randomness Arises in considering the extent to which a given infinite sequence (typically binary) can be seen as random. Broadly speaking, the issue is decided by putting the sequence through a number of suitable tests (essentially statistical and probabilistic); then sequences that pass the test may be called algorithmically random.

The field of such tests is large and varied, but e.g. a random sequence should not be readily compressible (in the sense of having a short encoding); it should be essentially unpredictable (so that no gambler can expect to win by betting on its unrevealed future); and it should pass all statistical tests for randomness (such as obeying the appropriate laws of large numbers). See Martin-Löf randomness. [R. Downey & D. Hirschfeldt, *Algorithmic Randomness and Complexity*. Springer, 2010]

aliasing (alias: **alias method; Walker's method**) A technique for sampling from the distribution $p = (p_1, \ldots, p_n)$ of a simple random variable X. It relies on the fact that X can be represented as a composite (or compound) random variable in the form $X = Y_U$, where U is uniform on $[1, \ldots, n-1]$, and $Y_k \in \{x_k, x_{a(k)}\}$, for a suitable function $a(\cdot) : \{1, \ldots, n-1\} \to [1, \ldots, n]$, called the alias function. Thus sampling from p is equivalent to sampling from a randomly chosen Bernoulli distribution. Put simply, one can always set

$$p = \frac{1}{n-1} \sum_{r=1}^{n-1} V_r,$$

where each V_r is a vector of length n with at most two non-zero entries, one of which can be forced to lie in the rth place of V_r by renumbering. [W. Hörmann *et al.*, *Automatic Nonuniform Random Variate Generation*. Springer, 2004]

Ali–Silvey divergence *see* f-divergence

Allais paradox (alias: **Allais phenomenon; Allais problem**) Arises in decision theory and utility theory. The Von Neumann–Morgenstern model for choice under uncertainty assumes that choices are independent of irrelevant alternatives. M. Allais constructed uncertain choices such that the irrelevant alternatives affected the choices of real decision makers. Behavioural decision theory accounts for this in a number of ways, including the certainty effect, or by postulating regret as a factor in decision-making. The paradox can be seen either as a failure of people to be rational, or as a failure of classical utility to account for people's behaviour. [I. Hacking, *Introduction to Probability and Inductive Logic*. Cambridge University Press, 2001]

Allison mixture Refers to the fact that two independent white noise processes can be mixed to yield a process that is not white noise. Formally, let $X_n^{(1)}$ and $X_n^{(2)}$ be independent stationary processes with means λ and $\mu, \lambda \neq \mu$, whose autocorrelation functions are both given by $\rho(X_n^{(i)}, X_{n+m}^{(i)}) = \delta(m)$, where $\delta(m)$ is the Kronecker delta. Let Z_n be a two-state Markov chain with transition matrix $\begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$, with $p + q \neq 1$, and $pq \neq 0$, in equilibrium. Then the process $W_n = X_n^{(Z_n)}$ is stationary, but is not white noise, in that $\rho(W_n, W_{n+m}) \neq \delta(m)$. [D. Abbott, Developments in Parrondo's paradox, in *Applications of Nonlinear Dynamics*, V. Longhini *et al.* (eds). Springer, 2009]

allocation problem (alias: **scheduling problem; sequencing problem**) A large class of problems in operations research that require the allocation of some resource (or vector of resources), which may be random, to satisfy various (possibly random) constraints, according to some criterion of optimality. For example: bin-packing; portfolio allocation, knapsack problems; bandwidth allocation; network routing; buffering; makespan scheduling; bandits; job-shop problems; repairman scheduling; etc. [S. Sarin *et al.*, *Stochastic Scheduling*, Cambridge University Press, 2010]

almost surely (alias: **with probability 1; almost everywhere; almost certainly; a.s.; a.e.**) An event A such that $\mathbb{P}(A) = 1$ is said to occur almost surely; by extension the same is said of any probability statement that holds on an event of probability 1. Thus e.g. for r.vs defined on the same space, to say that '$X = Y$, a.s.' is to say that $\mathbb{P}(\{\omega : X(\omega) = Y(\omega)\}) = 1$. Note that an almost sure event A need not be the certain event Ω; the complement of an almost sure event A is a **null event**. The term 'almost everywhere' is commonly used in the more general context of measure theory, where it is said of a property (or assertion) that holds everywhere except on a set of measure zero. [J. Stoyanov, *Counterexamples in Probability*, 3rd edn. Dover, 2013]

alpha-mixing (alias: **strong mixing**) A concept that formalizes the idea that a sequence $X_n, n \in \mathbb{Z}$, of random variables may not be independent, but widely separated members of the sequence may be negligibly dependent on each other. Formally, let A be any event that depends only on $\{X_r : r \leq k\}$, and B any event that depends only on $\{X_s : s \geq n + k\}$, and define

$$\alpha(n) = \sup_{k \geq 1} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

If, for any A and B, $\alpha(n) \to 0$ as $n \to \infty$, then the sequence X_n is said to be α-mixing. It is the weakest of a number of so-called strong-mixing conditions, of which independence is the strongest. [R. Bradley, Basic properties of strong mixing conditions, *Probab. Surv.*, 2, 107–144, 2005]

alternating renewal A random process that arises as a model for a system that works (or is on, or is up) for a random time to failure, and is then off (or down) for a random time until repair; this cycle is independently repeated ad infinitum. Formally, let $X_0, X_1, \ldots$ be a sequence of independent non-negative r.vs such that $X_{2r-1}, r \geq 1$ have distn $F(x)$, and $X_{2r}, r \geq 1$, have distn $G(x)$. This, together with the counting function $N(t) = \max\{n : \sum_{r=0}^{n} X_r \leq t\}$, comprises an alternating renewal process. It is ordinary if X_0 has the distn $G(x)$; otherwise it is delayed. [D. R. Cox, *Renewal Theory*. Methuen, 1962]

amalgamation paradox *see* Yule–Simpson paradox

ambit process Arises in modelling the evolution of suitable spatial processes in time. Broadly, for $t \in \mathbb{R}$ and $\mathbf{r} \in \mathbb{R}^d$, one defines a so-called ambit field as an integral $Y(t, \mathbf{r}) = \int k(s, t; \mathbf{x}, \mathbf{r}) M(\mathrm{d}s, \mathrm{d}\mathbf{x})$, where M is an appropriate random measure and k some (possibly random) kernel. Then the values of the field along a path in $\mathbb{R}^{d+1}$ constitute an ambit process. [O. Barndorff-Nielsen *et al.*, Ambit processes and SDES, in *Advanced Mathematical Methods for Finance*, G. Di Nunno & B. Øksendal (eds). Springer, 2011]

American option Arises in financial mathematics, where it refers to contracts in which the holder of an option can exercise it at any time up to the expiry (or maturity) date T. [M. Musiela & M. Rutkowski, *Martingale Methods in Financial Modelling*, 2nd edn. Springer, 2005]

Ammeter process Arises in insurance mathematics, where it denotes a Cox (i.e. doubly stochastic Poisson) process such that the random intensity $\lambda(t)$ is given for $t \geq 0$ by $\lambda(t) = A_n, n - 1 \leq t < n, n \geq 1$, where $A_n, n \geq 1$, is a sequence of i.i.d.r.vs. [J. Grandell, *Mixed Poisson Processes*. Chapman & Hall, 1997]

Andersen *see* Sparre Andersen (model, theorem, etc.)

Anderson's theorem (or inequality) In its simplest form this asserts that if X has an origin-symmetric unimodal density, and Y is independent of X, then $\mathbb{P}(|X| \le a) \ge \mathbb{P}(|X + Y| \le a)$. More generally, if the r.vr $\mathbf{X} \in \mathbb{R}^d$ has an origin-symmetric unimodal density, and A is an origin-symmetric convex set in $\mathbb{R}^d$, then $\mathbb{P}(\mathbf{X} \in A) \ge \mathbb{P}(\mathbf{X} + \mathbf{Y} \in A)$ for any $\mathbf{Y} \in \mathbb{R}^d$ that is independent of $\mathbf{X}$. [T. Anderson, The integral of a symmetric convex set and some probability inequalities, *Proc. Amer. Math. Soc.*, 6, 170–176, 1955]

annealing method (alias: **simulated annealing**) Arises in optimization. It is a probabilistic procedure for finding the global extremes of a function $f(x), x \in S$, that may have many widely dispersed local extremes in the state space S, which is commonly a lattice or a graph. In such cases, elementary optimization methods such as exhaustion, or gradient (hill-climbing) methods (possibly stochastic) are often useless. The method is a type of Markov chain Monte Carlo technique, and one version is implemented as follows.

Assume we seek a global minimum; simulate a Markov chain $X_n, n \ge 0$, on S, whose stationary distribution is $\pi(x) \propto \exp\{f(x)T^{-1}\}$ (being a Gibbs distn), where the arbitrary parameter T is called the temperature. As $T \downarrow 0$, this distribution is supported by the set M of global minima of $f(x)$. Therefore $T = T(n)$ is set as a non-increasing function of time, called the annealing schedule, and when T is small enough the chain is very likely to be in M. The nomenclature arises from the analogy with the creation of physical objects with low internal energy (and hence stress) by first heating and then cooling according to a good schedule. A related (mixture) method, in which a chain is run in equilibrium on a state space augmented by values of T, is called **simulated tempering**. *See* Gibbs sampling. [B. Gidas, Metropolis-Type Monte Carlo simulation algorithms and simulated annealing, in *Topics in Contemporary Probability and its Applications*, L. Snell (ed.). CRC Press, 1995]

annihilating particle system A type of interacting particle system in which particles may disappear on encountering another particle. An example is the annihilating random walk on $\mathbb{Z}^d$, in which all particles present at $t = 0$ perform independent simple random walks on the lattice, and any particles that meet at a vertex (or use the same edge), are instantly removed. Interest centres, e.g., on the asymptotic behaviour of the probability $p(t)$ that the origin (or other given site) is occupied at time t. More generally, particles may be of two (or more) types, with only different types (say) annihilated on meeting. Related processes include those where particles destroy themselves (death process), kill neighbours (coalescence), or reproduce (branching annihilation). [T. Liggett, *Interacting Particle Systems*. Springer, 2005]

announceable time *see* predictability

annuity Arises in actuarial maths. A periodic (classically, annual) payment, which may vary, and may run for a fixed term, or for the annuitant's lifetime, or forever (in which last case it may be called a perpetuity). Formally, the present value of an annuity (equal to the fair purchase price) is $V = \sum_{j=1}^{T} A_j \prod_{k=1}^{j-1} D_k$, where T is the duration of the annuity (lifetime), A_j are the amounts paid, and D_j are the discount factors subsuming the effects of inflation and interest rates. [D. Dickson *et al.*, *Actuarial Mathematics for Life Contingent Risks*. Cambridge University Press, 2009]

anomalous diffusion (alias: **enhanced diffusion, sub- and super diffusion**) Denotes a class of real-world processes that are essentially diffusive in nature, but whose detailed properties are different from those of the classic Brown–Einstein–Fick motion (BEF). If $X(t)$ is a BEF process, then its mean-square displacement $\mathbb{E}\big(X(t)^2\big)$ is proportional to the elapsed time t; in an anomalous diffusion $A(t), \mathbb{E}\big(A(t)^2\big) \propto t^a$, where $a < 1$, $a > 1$, and $a = 2$ correspond respectively to **subdiffusion**, **superdiffusion**, and **ballistic diffusion**. Examples include tracers in turbulence, transient currents in photocopiers, and diffusive percolation. A classic anomalous diffusion model is the Montroll–Weiss process, in which a particle makes i.i.d. random jumps in $\mathbb{R}^d$ at the instant of a renewal process. Another class of models comprises the fractional diffusion processes. These may be variously derived, but typically their probability distributions satisfy fractional diffusion equations, i.e. of the form $\partial^\alpha f / \partial t^\alpha = \Delta^\beta f$, where $\partial^\alpha / \partial t^\alpha$ is a time-fractional derivative, and Δ^β is a space-fractional differential operator. [D. Ben-Avraham & S. Havlin, *Diffusion and Reactions in Fractals and Disordered Systems*. Cambridge University Press, 2000]

Anscombe's condition For a sequence of r.vs $X_n, n \ge 0$, this condition amounts to requiring uniform continuity in probability. Formally, for any $\varepsilon > 0$ and $\delta > 0$, there exist $c > 0$ and $m < \infty$, such that for all $n > m$

$$\mathbb{P}\left(\max_{(1-c)n\le k\le(1+c)n}|X_k - X_n| > \varepsilon\right) < \delta.$$

[P. Berti *et al.*, An Anscombe-type theorem, *J. Math. Sci.*, 196(1), 15–22, 2014]

Anscombe's quartet Comprises four artificial sets of bivariate data (X, Y), having the same marginal means and variances, the same covariance, and the same regression of Y upon X. However, when plotted they exhibit entirely different apparent functional relationships between their respective X and Y. [F. Anscombe, Graphs in statistical analysis, *Amer. Stat.*, 27(1), 17–21, 1973]

Anscombe's theorem This is a limit theorem for randomly indexed sequences of random variables. Formally, let $X_n, n \ge 0$ satisfy Anscombe's condition and converge in distribution as $n \to \infty$; we write $X_n \to^D X$. If $M(n), n \ge 0$, is an integer-valued random process, and $a(n), n \ge 0$ a non-random sequence of positive numbers increasing to ∞ as $n \to \infty$, where $M(n)/a(n)$ converges in probability to 1, then $X_{M(n)}$ converges in distribution (with the same limit as X_n), as $n \to \infty$. Note that it is not assumed that $M(n)$ is independent of X_n. [A. Gut, Anscombe's theorem 60 years later, *Seq. Anal.*, 31(3), 368–396, 2012]

anticipating integral A type of stochastic integral in which the integral is not necessarily adapted to the σ-field generated by the integrator. That is to say, it fails to be predictable, i.e. non-anticipating. Examples include the **Skorokhod integral**, when the integrator is the Wiener process; and the **Ma–Protter–San Martin** integral, when the integrator is a normal martingale. [C. Tudor, Martingale-type stochastic calculus for anticipating integral processes, *Bernoulli*, 10(2), 313–325, 2004]

antithetic variables Arise in Monte Carlo methods, where they were introduced by J. Hammersley and K. Morton to improve efficiency by reducing the variance of outcomes. For example, if r.vs X and Y, with the same distn F, are simulated to estimate their common mean, the variance of the elementary estimator $\frac{1}{2}(X+Y)$ is $\frac{1}{2}\operatorname{var} X + \operatorname{cov}(X, Y)$. Independent simulations thus yield a larger variance than simulations such that $\operatorname{cov}(X+Y)<0$ (i.e. antithetic, from the Greek word meaning opposite). In this example, one may simulate a uniform r.v. U on $(0,1)$, and set $X = F^{-1}(U)$ and $Y = F^{-1}(1-U)$; it was shown by P. Moran that this minimizes the correlation between X and Y having the same distn. [S. Ross, *Simulation*, 5th edn. Academic Press, 2013]

a priori probability A probability determined independently of experience or empirical evidence (as it relates to the problem at hand). Such probabilities may be fixed subjectively, or by logical argument, or by applying some principle, such as insufficient reason or maximum entropy. Such a probability may be used as a prior in Bayes methods; but note that not all priors are determined a priori. [H. Jeffreys, *Theory of Probability*, 3rd edn. Oxford University Press, 1961]

arbitrage Arising in financial mathematics, this denotes any collection of financial transactions, or betting scheme, that yields a gain with probability 1. (A weak arbitrage yields no loss with probability 1, and a positive probability of positive gain.) A set of offered bets, or offered prices for assets and derivatives, is arbitrage-free (or a no-arbitrage market), if no such collection (or portfolio) exists. Standard market models for pricing derivatives and options require that the result must be arbitrage-free; in the simpler models this is sufficient to determine a unique fair price. More general models may require other (stronger) principles to constrain uniqueness of prices. A set of betting odds that offers an arbitrage opportunity is called a **Dutch book**. [T. Björk, *Arbitrage Theory in Continuous Time*, 3rd edn. Oxford University Press, 2009]

arbitrage theorem (alias: **no-arbitrage theorem; first fundamental theorem of finance**) Arises in the context of betting, with natural applications to financial mathematics. Informally, suppose that n betting opportunities are available on a game (or experiment) with m possible outcomes, and the amounts to be bet on each opportunity are free to choose. Then the theorem asserts that either there is a probability distribution on the outcomes such that all bets are fair (in the sense of expected return), or there is a betting scheme which wins with probability 1. Formally, let the betting scheme (amounts bet on each of the n opportunities) be $(b_1, \ldots, b_n) \in \mathbb{R}^n$, with the total return, when the outcome of the game is j, denoted by $r(\mathbf{b}, j) = \sum_{k=1}^n b_k r(k,j)$, where $r(k,j)$ is the return on b_k for outcome j. Then either there is a probability distn $(p_1, \ldots, p_m)$ such that, for all k, $\sum_{j=1}^m p_j r(k,j) = 0$, or there is a betting scheme $(a_1, \ldots, a_n)$ such that, for all j, $\sum_{k=1}^n a_k r(k,j) > 0$; yielding an arbitrage. This theorem may be derived from the duality theorem of linear programming. [S. Le Roy & J. Werner, *Principles of Financial Economics*. Cambridge University Press, 2001]

ARCH process An acronym for autoregressive conditionally heteroscedastic process; the word