International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

RANDOMIZED MACHINES: A FRAMEWORK FOR PROBABILISTIC INPUT-OUTPUT SYSTEMS

Kamran Alam Khan,
Department of Mathematics, V.R.A.L. Govt. Mahila Mahavidyalaya, Bareilly (U.P.)
E-mail: kamran12341@gmail.com

Abstract. The Markov kernel is a well-established mathematical object in probability theory. It provides a way to describe conditional probability distributions on measurable spaces. Although Markov kernels are widely used in many areas such as communication theory, randomized computation, stochastic control, game theory, machine learning, and cognitive science, these different systems are rarely described within one common operational framework that makes their shared structure clear and usable across fields. We introduce *Randomized Machines* as precisely this framework. A Randomized Machine $M=(X,Y,K)$ is an operational device for systems and situations with random outputs, where X is an input space, Y an output space, and K a Markov kernel assigning to each input $x\in X$ a probability distribution $K(x, .)$ on Y. In this construction, the device and its executions are taken as primitive, and the kernel emerges as their statistical summary. We develop the framework at three levels of generality (finite, discrete, general) and discuss the theory of General Randomized Machines (GRMs). We then show that many heterogeneous domains such as noisy communication channel, randomized algorithm, probabilistic Turing machine, Markov chains, Markov decision process, mixed strategies and game dynamics, probabilistic classifiers, language models etc are all instances of a single triple $M=(X,Y,K)$.

Keywords: randomized machines; Markov kernels; general randomized machines; Giry monad; Kleisli morphisms; communication theory; noisy communication channel; randomized computation; randomized algorithm; probabilistic Turing machine; stochastic control; Markov chains; Markov decision process; game theory; mixed strategies and game dynamics; decision theory; machine learning; probabilistic classifiers; language models; and cognitive science.

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

2020 Mathematics Subject Classification: 60J05, 18C20, 94A40, 68Q87, 90C40, 60J10, 91A05

1. Introduction

Many systems across mathematics, computer science, economics, and the natural sciences share a common operative structure where a deterministic input is supplied to a system, and the system produces a probabilistic output governed by a distribution that depends on the input. A chess player at a given position selecting a move randomly according to a mixed strategy; A noisy communication channel flips a transmitted bit with known conditional probabilities ([19]); the conditional probability table of a Bayesian network tells us how a child variable is distributed given its parents; a randomized algorithm whose output distribution depends on its input ([16]), a learning agent selecting actions according to a stochastic policy ([9]), a probabilistic classifier outputting a distribution over class labels conditioned on the feature vector ([8]); a language model assigns a distribution to the next token given the text so far; a quantum measurement returns one of several outcomes with probabilities fixed by the Born rule; a stochastic policy chooses an action at random given the current observation and a cognitive organism responding to stimuli with learned probabilities ([3]); each of these is, at its mathematical core, a mapping from a deterministic input to a probability distribution over outputs.

The underlying mathematical object common to all these systems is a *Markov kernel*, a standard tool in probability theory ([1], [2]) and categorical probability ([6], [7]). Yet these systems are rarely gathered under a single unified operational framework that emphasizes their shared character as input-output devices. The purpose of this paper is to introduce Randomized Machines as precisely such a framework. A Randomized Machine $M=(X,Y,K)$ is an operational device such that when supplied with input $x\in X$, it produces a single definite output in Y drawn from the distribution $K(x, .)$. The kernel K is the mathematical specification and the machine is the operational entity. The conceptual distinction is significant as in stochastic process theory a Markov kernel is an abstract primitive from which processes are constructed and in our framework the device and its executions are primitive, and K is the statistical summary of their behavior. We do not claim to enrich the theory of Markov kernels or to extend categorical probability. Rather,

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

we propose an operational framework using Markov kernels with some concrete conceptual value. We define and study different randomized machines. We then show that heterogeneous domains such as noisy communication channels, probabilistic automata, Markov decision processes, mixed game-theoretic strategies, probabilistic classifiers, and stimulus-response models in cognitive science are all instances of a single triple $M=(X,Y,K)$. Domain-specific results that previously required separate derivations become corollaries of a common framework.

2. Preliminaries

Throughout, a *measurable space* is a pair (S,Σ_S) and a *standard Borel space* is a measurable space isomorphic to a Borel subset of $\mathbb{R}$ with its Borel σ-algebra. Finite and countable sets with the power-set σ-algebra, $\mathbb{R}^n$, and countable products of such spaces are standard Borel. For measure-theoretic background we use ([1], [18]).

Definition 2.1 *(Markov kernel).* Let (S_1,Σ_1) and (S_2,Σ_2) be measurable spaces. A *Markov kernel* (also called a stochastic kernel or Markov kernel) from S_1 to S_2 is a function $K:S_1\times\Sigma_2\to[0,1]$ such that

(i) for every $s\in S_1$, the set function $K(s,\cdot)$ is a probability measure on (S_2,Σ_2);

(ii) for every $E\in\Sigma_2$, the map $s\mapsto K(s,E)$ is Σ_1-measurable.

We write $K:S_1\rightsquigarrow S_2$ for the Markov kernel from S_1 to S_2. Condition (i) says that each s selects a probability distribution; condition (ii) says that this selection depends measurably on s, which is exactly what is needed to integrate against the kernel. For a probability measure μ on S_1 and a bounded measurable $\varphi:S_2\to\mathbb{R}$ we write

$$(\mu K)(E)=\int_{S_1} K(s,E)\,\mu(ds), \qquad (K\varphi)(s)=\int_{S_2} \varphi(t)\,K(s,dt). \tag{2.1}$$

By a monotone-class argument $K\varphi$ is bounded and measurable, μK is a probability measure, and $\int\varphi\,d(\mu K)=\int K\varphi\,d\mu$. The *composition* of a kernel K_1 from S_1 to S_2 with a kernel K_2 from S_2 to S_3 is

$$(K_2\circ K_1)(s,E)=\int_{S_2} K_2(t,E)\,K_1(s,dt). \tag{2.2}$$

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Lemma 2.1. The composition $K_2 \circ K_1$ is a Markov kernel from S_1 to S_3, and $(K_2 \circ K_1)\varphi = K_1(K_2\varphi)$ for every bounded measurable φ.

Proof. For fixed s, countable additivity of $E \mapsto \int K_2(t,E)K_1(s,dt)$ follows from the monotone convergence theorem, and $(K_2 \circ K_1)(s,S_3) = 1$. Measurability in s holds because $t \mapsto K_2(t,E)$ is bounded and measurable and K_1 is a kernel. The identity for functions holds for indicators by definition, for simple functions by linearity, and for bounded measurable functions by uniform approximation. ∎

The *total-variation distance* of two probability measures P, Q on (S, Σ_S) is

$$\mathrm{TV}(P,Q) = \sup_{E \in \Sigma_S} |P(E) - Q(E)| = \sup_{0 \le f \le 1} \left| \int f\,dP - \int f\,dQ \right|. \tag{2.3}$$

On a finite set, $\mathrm{TV}(P,Q) = 1/2 \sum_y |P(y) - Q(y)|$. If φ is bounded and measurable with oscillation $\mathrm{osc}(\varphi) = \sup \varphi - \inf \varphi$, then $\left| \int \varphi\, dP - \int \varphi\, dQ \right| \le \mathrm{osc}(\varphi)\, \mathrm{TV}(P,Q)$. The Shannon entropy of a distribution P on a finite set is $H(P) = -\sum_y P(y) \log_2 P(y)$, and the relative entropy is $D(P \| Q) = \sum_y P(y) \log_2 \frac{P(y)}{Q(y)}$ [19].

3. Randomized machines

Definition and basic examples

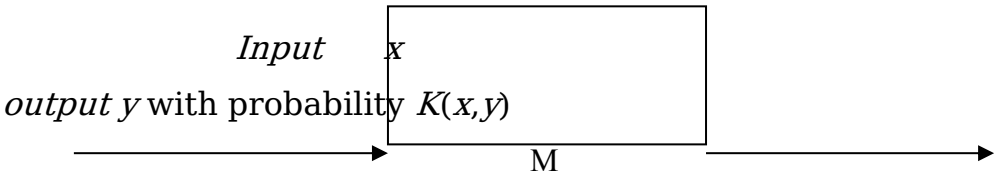
We now propose our operational framework using Markov kernels to define Randomized machines:

Definition 3.1 (Finite Randomized Machine): Let $X = \{x_1, x_2, \ldots x_m\}$ and $Y = \{y_1, y_2, \ldots y_n\}$ be non-empty finite sets, called the *input space* and *output space* respectively. A *Finite Randomized Machine* is a triple $M = (X, Y, K)$, where the kernel $K: X \times Y \to [0,1]$ satisfies $K(x,y) \ge 0 \quad \forall x \in X, y \in Y$ and $\sum_{y \in Y} K(x,y) = 1 \quad \forall x \in X$. The kernel K is equivalently an $m \times n$ row-stochastic matrix $\left(K(x_i, y_j)\right)$.

When input $x \in X$ is supplied, the machine M produces a random output $y \in Y$ such that the probability

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

$$\mathbb{P}(M(x)=y)=K(x,y)$$



The output in any individual execution is a definite element of Y; the kernel $K(x,.)$ is the probability law governing which element is produced. We write $M(x)\sim K(x,.)$. Therefore a Randomized Machine is not a function from X to Y but a probabilistic rule. Also it is not just a kernel but a realized process that samples outputs according to the kernel.

Example 3.1.1 Let $X=\{x_a,x_b\}$ and $Y=\{y_1,y_2,y_3\}$ and define the triple $M=(X,Y,K)$ by $K(x_a,\cdot)=(0.6, 0.3, 0.1)$ and $K(x_b,\cdot)=(0.1, 0.9, 0)$. On input x_a the machine returns y_1,y_2,y_3 with probabilities $0.6,0.3,0.1$; on input x_b with probabilities $0.1,0.9,0$. Every row sums to one, so this is a Finite Randomized Machine. Its matrix is

$$\mathrm{K}_M=\begin{pmatrix}0.6 & 0.3 & 0.1\\ 0.1 & 0.9 & 0\end{pmatrix}. \qquad (3.1)$$

The operation of this machine can be illustrated by noting, for example, that the machine M has 60% chance to produce an output 'y_1' when input 'x_a' is given. Also there is no possibility to obtain output 'y_3' when input 'x_b' is given.

Example 3.1.2 (Noisy Sensor) Let $X=\{Low, High\}$ be true states and $Y=\{Low, High\}$ be reported states. Suppose a sensor M reports the true state with 90% accuracy. Then the kernel K associated with M is given by $K(Low, Low)=0.9$, $K(Low, High)=0.1$, $K(High, Low)=0.1$ and $K(High, High)=0.9$. The sensor M can be formulated as a Randomized Machine such that

$$M(Low)=\{(Low,0.9),(High,0.1)\}$$
$$M(High)=\{(Low,0.1),(High,0.9)\}$$

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Example 3.1.3 (Uniform Randomizer) Let an unbiased dice be thrown. Then the input space and output space can be given as follows:

$$X=\{Roll\},\quad Y=\{1,2,3,4,5,6\}$$

If we define the kernel K by $K(Roll, y)=\frac{1}{6}\ \forall y\in Y$, then the Randomized Machine $M=(X,Y,K)$ simulates the roll of an unbiased dice and $M(x)\sim Uniform(\{1,2,3,4,5,6\})$.

Example 3.1.4 (Binary symmetric channel) Let $X=Y=\{0,1\}$ and $\varepsilon\in\left[0,\frac{1}{2}\right)$. Define

$$K(x,y)=\begin{cases}1-\varepsilon & if\ y=x\\ \varepsilon & if\ y\neq x\end{cases}$$

The 2×2 stochastic matrix is $K=\begin{pmatrix}1-\varepsilon & \varepsilon\\ \varepsilon & 1-\varepsilon\end{pmatrix}$. Each row sums to 1. The machine $M=(X,Y,K)$ transmits bit x and receives a correct copy with probability $1-\varepsilon$ and an erroneous copy with probability ε.

Example 3.1.5 (LLM Prediction) The LLM's next-token prediction mechanism can be represented as a finite randomized machine $M=(X,Y,K)$, where X is the set of finite token sequences representing input prefixes, and Y is the finite vocabulary of tokens of a large language model (LLM) and whose stochastic behavior is specified by the Markov kernel K. For example,

M ("*Honesty is the*") ={

(*best*,0.512356896521235), (*key*,0.152956117527256),

(*virtue*,0.222356954232257),(*bad*,0.023987361155349),... }

Here probabilities are taken arbitrarily to illustrate the operation of LLM as a finite randomized machine.

Definition 3.2 (Discrete Randomized Machine): Let X and Y be non-empty countable sets, called the *input space* and *output space* respectively. A *Discrete Randomized Machine* is a triple $M=(X,Y,K)$, where the kernel $K:X\times Y\to[0,1]$ satisfies $K(x,y)\geq0$

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

$\forall x\in X, y\in Y$ and $\sum_{y\in Y} K(x,y) = 1 \ \forall x\in X$, the sum being a countably infinite series when Y is infinite.

Every Finite Randomized Machine is a discrete Randomized Machine.

Example 3.2.1 Let $X=\{1,2,3,\ldots\}$ and $Y=\{0,1,2,\ldots\}$ both countably infinite sets. For each $x\in X$, define $p_x := 1/(x+1) \in (0,1)$ and set $K(x,y) := (1-p_x)^y p_x$, $y\in Y$. Then $K(x,y)\geq 0$ for all x, y and

$$\sum_{y=0}^{\infty} K(x,y) = p_x \sum_{y=0}^{\infty} (1-p_x)^y = \frac{p_x}{1-(1-p_x)} = 1$$

Thus $M=(X,Y,K)$ is a Discrete Randomized Machine. When input x is supplied, the output $M(x)$ counts the number of failures before the first success in a sequence of independent Bernoulli trials each with success probability $p_x := 1/(x+1)$. The machine therefore maps a positive integer x to a random non-negative integer drawn from the Geometric $(1/(x+1))$ distribution.

Definition 3.3 *(General Randomized Machine).* Let (X,Σ_X) be a measurable *input space* and (Y,Σ_Y) a measurable *output space*. A *General Randomized Machine* (GRM) is a triple $M=(X,Y,K)$, where $K: X\times\Sigma_Y \to [0,1]$ is a Markov kernel, called the *characteristic kernel* of M. When an input $x\in X$ is supplied, the machine produces a random output $M(x)$ of Y satisfying

$$\mathbb{P}(M(x)\in E) = K(x,E), \qquad E\in\Sigma_Y. \tag{3.2}$$

Thus given an input $x\in X$, the General Randomized Machine M produces a random output $y\in Y$ whose distribution is given by the probability measure $K(x, \ .)$.

Two features of this definition are worth stressing. First, nothing persists between two evaluations: if the machine is run twice on the same input, the two outputs are independent draws. Second, the kernel is the *complete* specification of the machine; any further structure (a program, a neural network, a physical device) matters only through the kernel it realizes.

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Example 3.3.1 (Gaussian channel) Let $X=Y=\mathbb{R}$ with Borel σ-algebra. Let $\rho>0$, $B\in B(\mathbb{R})$. Define

$$K(x,B)=\int_B \frac{1}{\sqrt{2\pi\rho^2}}\exp\left(-\frac{(y-x)^2}{2\rho^2}\right)dy$$

Then $K(x,.)=N(x,\rho^2)$, i.e. the input x produces a Gaussian-noise-corrupted output. This GRM adds independent Gaussian noise of variance ρ^2 to its input. It is a continuous-space machine whose kernel is given by the density $k(x,y)=(2\pi\rho^2)^{-1/2}\exp\ (-(y-x)^2/2\rho^2)$.

Example 3.3.2. A mixed strategy of a player in a game is a GRM from the player's information set (or from a trivial one-point set) to the set of actions ([15]). Its kernel is the probability vector over actions.

In the finite case the matrix K_M with $(\mathrm{K}_M)_{ij}=K(x_i,\{y_j\})$ is row-stochastic. Pushing a prior row vector μ through the machine gives the output law $\mu\mathrm{K}_M$, and applying the machine to a column vector φ of payoffs gives the vector of expected payoffs $\mathrm{K}_M\varphi$; these are the matrix forms of the two actions in (2.1). This matrix view is standard for Markov chains and channels ([13]).

Remark 3.1 Every Discrete (and therefore every Finite) Randomized Machine is a General Randomized Machine; the above three definitions form a strict hierarchy.

Theorem 3.1 (Equivalence) Let (X,Σ_X) and (Y,Σ_Y) be measurable spaces. The correspondence $M=(X,Y,K)\longmapsto K$ is a bijection between the class of General Randomized Machines on (X,Σ_X) and (Y,Σ_Y), and the class of Markov kernels from X to Y.

Proof The result directly follows from definition 3.3.

Remark 3.2 (Conceptual distinction) Despite the bijection established in Theorem 3.1, the two frameworks are conceptually distinct:

(i) In stochastic process theory, K is a primitive from which processes are built via the Kolmogorov extension theorem. In the Randomized Machine framework, the device and its executions are primitive; i.e. $\mathbb{P}(M(x)\in B)=K(x,B)$ is derived as a statistical summary.

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

(ii) A kernel specifies a measure but not how a sample is generated. A machine has an explicit mechanism (mixed strategy, probabilistic circuit, random tape etc) producing a definite output on each run.

Representation theorems

A GRM is specified only by its kernel, but it is useful to know how a kernel can be realized by drawing a random variable. The following theorem says that one uniform random number is always enough. It is a form of the randomization lemma of probability theory [1].

Theorem 3.2 *(Seed representation).* Let (X,Σ_X) be a measurable space, (Y,Σ_Y) a standard Borel space and $M=(X,Y,K)$ a GRM. Then there is a measurable map $\widetilde{M}:X\times[0,1]\to Y$ such that for every $x\in X$ and $E\in\Sigma_Y$,

$$K(x,E)=\lambda\left(\{r\in[0,1]:\widetilde{M}(x,r)\in E\}\right), \tag{3.3}$$

where λ is Lebesgue measure. Equivalently, if $U\sim\text{Uniform}[0,1]$, then $\widetilde{M}(x,U)\sim K(x,\cdot)$.

Proof *Case* $Y=\mathbb{R}$. Let $G(x,y)=K(x,(-\infty,y])$. For each y the map $x\mapsto G(x,y)$ is measurable, and for each x the map $y\mapsto G(x,y)$ is a distribution function. Define the quantile function $Q(x,r)=\inf\{y\in\mathbb{Q}:G(x,y)\geq r\}$ for $r\in(0,1)$ and $Q(x,r)=0$ for $r\in\{0,1\}$. By right-continuity of $G(x,\cdot)$, for $r\in(0,1)$ and real y_0 we have $Q(x,r)\leq y_0$ if and only if $r\leq G(x,y_0)$. Hence $\{(x,r):Q(x,r)\leq y_0\}=\{(x,r):r\leq G(x,y_0)\}\cup\{(x,r):r\in\{0,1\},0\leq y_0\}$ is measurable, so Q is measurable, and $\lambda(Q(x,\cdot)\leq y_0)=G(x,y_0)$.

General case. Let $\phi:Y\to B$ be a Borel isomorphism onto a Borel set $B\subseteq\mathbb{R}$. The kernel $\widetilde{K}(x,\cdot)=K\left(x,\phi^{-1}(\cdot)\right)$ on $\mathbb{R}$ is concentrated on B. Let $\widetilde{Q}$ be its quantile map and fix $y_0\in Y$. Put $\widetilde{M}(x,r)=\phi^{-1}\left(\widetilde{Q}(x,r)\right)$ if $\widetilde{Q}(x,r)\in B$ and $\widetilde{M}(x,r)=y_0$ otherwise. Then $\widetilde{M}$ is measurable, and for each x we have $\widetilde{Q}(x,U)\in B$ almost surely, so $\widetilde{M}(x,U)=\phi^{-1}\left(\widetilde{Q}(x,U)\right)$ has law $K(x,\cdot)$. ∎

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Note 3.1: For finite $Y=\{y_1,\ldots,y_m\}$ the map $\widetilde{M}$ is the familiar partition of $[0,1]$ into intervals of lengths $K(x,y_j)$.

Behavioral equivalence and Canonical reduction

The operational definition of a Randomized Machine leads naturally to a precise notion of *when two machines are the same*.

Definition 3.4 (Behavioral equivalence) Two Randomized machines $M_1=(X,Y,K_1)$ and $M_2=(X,Y,K_2)$ are said to be **behaviorally equivalent**, **written** $M_1\sim M_2$ if $K_1(x,B)=K_2(x,B)$ for all $x\in X$ and $B\in\Sigma_Y$.

Definition 3.5 Let $M=(X,Y,K)$ be a randomized machine. Two inputs $x_1,x_2\in X$ are said to be **operationally indistinguishable**, written $x_1\simeq x_2$ if $K(x_1,.)=K(x_2,.)$ as probability measures on Y.

Theorem 3.3 (Canonical reduction) The relation $\simeq$ is an equivalence relation on X. The *reduced machine* $\widetilde{M}=(X/\simeq,Y,\widetilde{K})$, where $\widetilde{K}([x],B):=K(x,B)$, is well-defined, behaviorally equivalent to M, and is the unique machine (up to bijection on the input space) with the input space among all machines behaviorally equivalent to M.

Proof Reflexivity, symmetry, and transitivity are immediate. For the well-definedness, we observe that $x_1\simeq x_2$ if $K(x_1,.)=K(x_2,.)$, by definition 3.5, i.e. $\widetilde{K}([x_1],B)=\widetilde{K}([x_2],B)$ for all B. The behavior equivalence $\widetilde{M}\sim M$ follows immediately. For minimality, suppose there exists a machine $M'=(X',Y',K')$ equivalent to $M=(X,Y,K)$. Then we can define a surjection $f:X'\to X/\simeq$, such that $K'(x',B)=K(f(x'),B)$, therefore $|X'|\geq|X/\simeq|$.

4. Algebra of general randomized machines

4.1 Sequential composition

Proposition 4.1 Let $M_1=(X,Y,K_1)$ and $M_2=(Y,Z,K_2)$ be two General Randomized Machines. Then the triple (X,Z,K_2oK_1) is also a General Randomized Machine.

Proof We verify that K_2oK_1 is a Markov kernel from X to Z. Let $K:=K_2oK_1$.

(i) $K(x,.)$ *is a probability measure on* (Z,Σ_Z)*:*

Non-negativity: Since $K_2(y,C)\geq 0$ for all y and C, we have

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

$$K(x,C)=(K_2 oK_1)(x,C)=\int_Y K_2(y,C)\,K_1(x,dy)\geq 0$$

Normalization: We have

$$K(x,Z)=\int_Y K_2(y,Z)\,K_1(x,dy)=\int_Y 1\,K_1(x,dy)=1$$

Since $K_2(y,Z)=1$ for all y.

Countable additivity: Let $\langle C_n\rangle_{n\geq 1}\subset\Sigma_Z$ be pairwise disjoint. Since $0\leq K_2(y,C)\leq 1$, the monotone convergence theorem gives

$$\begin{aligned}K\left(x,\bigcup_n C_n\right)&=\int_Y K_2\left(y,\bigcup_n C_n\right)K_1(x,dy)\\&=\int_Y\sum_n K_2(y,C_n)\;K_1(x,dy)\\&=\sum_n\int_Y K_2(y,C_n)\,K_1(x,dy)\\&=\sum_n K(x,C_n)\end{aligned}$$

(ii) $x\longmapsto K(x,C)$ *is* Σ_X*-measurable for each fixed* $C\in\Sigma_Z$

For each fixed $C\in\Sigma_Z$, the function $y\longmapsto K_2(y,C)$ is Σ_Y-measurable. Therefore the map $x\longmapsto K(x,C)=\int_Y K_2(y,C)\,K_1(x,dy)$ is Σ_X-measurable.■

For finite spaces with stochastic matrices K_1 and K_2, the matrix of $M_2\circ M_1$ is the product $\mathrm{K}_1\mathrm{K}_2$.

In light of above theorem, we have the following definition:

Definition 4.1 (Sequential Composition of Machines) The *sequential composition* of the machines $M_1=(X,Y,K_1)$ and $M_2=(Y,Z,K_2)$ is the machine $M_2 oM_1$, given by

$$M_2 oM_1=(X,Z,K_2 oK_1)$$

Operationally $M_2 oM_1$ runs on input $x\in X$, first giving $y\sim K_1(x,\cdot)$ and then giving output $z\sim K_2(y,\cdot)$, using fresh randomness at the second step, that is, $z\sim(K_2 oK_1)(x,.)$.

Example 4.1.1. Let $M=(\{x_a,x_b\},\{y_1,y_2,y_3\},K_M)$ where $K_M(x_a,\cdot)=(0.6,0.3,0.1)$ and $K_M(x_b,\cdot)=(0.1,0.2,0.7)$. Let $N=(\{y_1,y_2,y_3\},\{w_1,w_2\},K_N)$ where $K_N(y_1,\cdot)=(0.8,0.2)$, $K_N(y_2,\cdot)=(0.3,0.7)$ and $K_N(y_3,\cdot)=(0.1,0.9)$. Then the sequential composition of the machines M and N is the machine $NoM=(\{x_a,x_b\},\{w_1,w_2\},K_N oK_M)$, where

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

$$K_{N\circ M}=\begin{pmatrix}0.6 & 0.3 & 0.1\\ 0.1 & 0.2 & 0.7\end{pmatrix}\begin{pmatrix}0.8 & 0.2\\ 0.3 & 0.7\\ 0.1 & 0.9\end{pmatrix}=\begin{pmatrix}0.58 & 0.42\\ 0.21 & 0.79\end{pmatrix}.$$

Definition 4.2 (*The identity machine)* Let (X,Σ_X) be a measurable space. The machine $I_X=(X,X,\delta)$ with kernel $\delta(x,A)=1_A(x)$, i.e. $\delta(x,\cdot)=\delta_x$ is called the identity machine.

Proposition 4.2 For every General Randomized Machine $M=(X,Y,K)$, we have $M\circ I_X=M=I_Y\circ M$.

Proof. Let $M=(X,Y,K)$ be any General Randomized Machine. Fix $x\in X$ and $E\in\Sigma_Y$. We have

$$(K\circ\delta)(x,E)=\int_X K(x',E)\,\delta(x,dx')=\int_X K(x',E)\,\delta_x(dx')=K(x,E)$$

Since integrating against the Dirac measure δ_x evaluates the integrand at x.

Now for fixed $x\in X$ and $E\in\Sigma_Y$, we have

$$(\delta\circ K)(x,E)=\int_Y \delta(y',E)\,K(x,dy')=\int_Y 1_E(y')\,K(x,dy')=K(x,E)$$

Therefore, $M\circ I_X=M=I_Y\circ M$ ∎

Proposition 4.3 *(Associativity).* For GRMs $M_1=(X,Y,K_1)$, $M_2=(Y,Z,K_2)$ and $M_3=(Z,W,K_3)$,

$$(M_3\circ M_2)\circ M_1=M_3\circ(M_2\circ M_1). \qquad (4.2)$$

Proof Fix $x\in X$ and $D\in\Sigma_W$. Applying the composition formula twice, we have

$$[K_3\circ(K_2\circ K_1)](x,D)=\int_Z K_3(z,D)\,(K_2\circ K_1)(x,dz)$$
$$=\int_Z K_3(z,D)\left[\int_Y K_2(y,dz)\,K_1(x,dy)\right] \text{ ..... (4.3)}$$

Also we have

$$[(K_3\circ K_2)\circ K_1](x,D)=\int_Y (K_3\circ K_2)(y,D)\,K_1(x,dy)$$
$$=\int_Y\left[\int_Z K_3(z,D)\,K_2(y,dz)\right]K_1(x,dy) \text{ ..... (4.4)}$$

Define the function $f(y,z):=K_3(z,D)$, $(y,z)\in Y\times Z$. Since $K_3(.,D)$ is Σ_Z-measurable, the function f is measurable with respect to $\Sigma_Y\otimes\Sigma_Z$. Moreover, $0\le f(y,z)\le 1$ for all (y,z), so f is bounded and non-negative. Since $K_3\circ K_2$ is a kernel, for each fixed $y\in Y$, $\int_Z f(y,z)\,K_2(y,dz)=(K_3\circ K_2)(y,D)\in[0,1]$ is Σ_Y-measurable in y.

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

By Fubini-Tonelli theorem applied to the product measure on $Y\times Z$ with σ-finite factors $K_1(x,.)$ and $K_2(y,.)$ the iterated integrals in (4.3) and (4.4) are equal. Therefore the two kernels $K_3 o(K_2 oK_1)$ and $(K_3 oK_2)oK_1$ agree on every measurable set $D\in\Sigma_W$. Hence we conclude that $M_3 o(M_2 oM_1)=(M_3 oM_2)oM_1$ ■

Definition 4.3 *(Deterministic GRM).* A GRM $M=(X,Y,K)$ is *deterministic* if $K(x,\cdot)$ is a Dirac measure for every x.

Proposition 4.4 *(Characterization of deterministic machines).* Let Y be standard Borel. A GRM $M=(X,Y,K)$ is deterministic if and only if $K(x,E)\in\{0,1\}$ for all $x\in X$ and $E\in\Sigma_Y$. In that case $K(x,\cdot)=\delta_{f(x)}$ for a measurable function $f:X\to Y$, i.e. $M=M_f$.

Proof A Dirac measure is $\{0,1\}$-valued. Conversely, assume $K(x,\cdot)$ is $\{0,1\}$-valued and let $\{E_n\}$ be a countable family generating Σ_Y and separating points (it exists because Y is standard Borel). The set $A_x=\bigcap\{E_n:K(x,E_n)=1\}\cap\bigcap\{E_n^c:K(x,E_n)=0\}$ is a countable intersection of sets of measure one, so $K(x,A_x)=1$ and $A_x\neq\emptyset$. Since the family separates points, A_x is a single point $f(x)$ and $K(x,\cdot)=\delta_{f(x)}$. Finally, $\{x:f(x)\in E\}=\{x:K(x,E)=1\}$ is measurable. ∎

Thus every measurable function $f:X\to Y$ defines the *deterministic GRM* $M_f=(X,Y,K_f)$ with kernel $K_f(x,E)=\delta_{f(x)}(E)=\mathbf{1}_E(f(x))$. In particular the *identity machine* $I_X=(X,X,\delta)$ has kernel $\delta(x,A)=\mathbf{1}_A(x)$, i.e. $\delta(x,\cdot)=\delta_x$.

Definition 4.4 (Feedback network) For the machine $M=(X,Y,K)$ with $X=Y$, i.e. the machine $M=(X,X,K)$, the *feedback network* from initial state $\xi_0=x_0\in X$ is the sequence $\langle\xi_n\rangle_{n\geq 0}$ defined by $\xi_{n+1}\sim K(\xi_n,.)$ independently of $\xi_0,\xi_1,....\xi_{n-1}$ given ξ_n.

The following is the central conceptual theorem of the paper. Its mathematical content is classical. The novelty of the representation lies in expressing it in Randomized machine framework.

Theorem 4.1 (Markov Property) Let $M=(X,X,K)$ be a General randomized machine and let $\langle\xi_n\rangle_{n\geq 0}$ be its feedback network from $\xi_0=x_0\in X$. Then

(i) $\langle\xi_n\rangle_{n\geq 0}$ is a time-homogeneous Markov chain on (X,Σ_X) with transition kernel K:

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

$$\mathbb{P}(\xi_{n+1}\in B|\xi_0,\xi_1,....,\xi_n)=K(\xi_n,B) \qquad \forall B\in\Sigma_X,\ n\geq 0$$

(ii) The n-step transition kernel is $K^n := KoKo...oK\ (n\ times)$, and for every $B\in\Sigma_X$,

$$K^n(x,B)=\mathbb{P}(\xi_n\in B|\xi_0=x)$$

and if X is finite with transition kernel matrix K, then K^n corresponds to the matrix power K^n, recovering the Chapman-Kolmogorov equation.

Proof (i) By the definition of the feedback network, ξ_{n+1} is drawn from $K(\xi_n,.)$ independently of $\xi_0,\xi_1,....,\xi_{n-1}$ given ξ_n. Therefore, for any $B\in\Sigma_X$ and any realization of $x_0,....,x_n$ of $(\xi_0,....,\xi_n)$, we have

$$\mathbb{P}(\xi_{n+1}\in B|\xi_0=x_0,....,\xi_n=x_n)=\mathbb{P}(M(x_n)\in B)=K(x_n,B)$$

The right-hand side depends only on $\xi_n=x_n$, not on the earlier values $x_0,....,x_n$. This is the Markov property. Time homogeneity holds because the kernel K is the same at every step.

(ii) We proceed by induction on n. The base case is the statement

$K^1(x,B)=\mathbb{P}(\xi_1\in B|\xi_0=x)$ which holds by definition.

Assume $K^n(x,B)=\mathbb{P}(\xi_n\in B|\xi_0=x)$ for all x and B. Then

$$\begin{aligned}\mathbb{P}(\xi_{n+1}\in B|\xi_0=x)&=\int_X\mathbb{P}(\xi_{n+1}\in B|\xi_n=x')\,\mathbb{P}(\xi_n\in dx'|\xi_0=x)\\&=\int_X K(x',B)\,K^n(x,dx')\\&=(KoK^n)(x,B)=K^{n+1}(x,B)\end{aligned}$$

In finite case, let $X=\{s_1,....,s_N\}$, the kernel composition formula

$K^{n+1}(s_i,\{s_j\})=\sum_{k=1}^{N}K^n(s_i,\{s_k\})\,K(s_k,\{s_j\})$ is exactly the (i,j)-entry of the matrix product $K^nK=K^{n+1}$, which is the Chapman-Kolmogorov equation.

4.2 Categorical structure

Theorem 4.2 General Randomized Machines with sequential composition and the identity machine $I_X=(X,X,\delta)$ where $\delta(x,A) := 1_A(x)$, form a category.

Proof (i) The sequential composition of the machines $M_1=(X,Y,K_1)$ and $M_2=(Y,Z,K_2)$ is the machine M_2oM_1, given by

$$M_2oM_1=(X,Z,K_2oK_1\ )$$

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Where the kernel $K_2 o K_1 : X \rightsquigarrow Z$ is defined by

$$(K_2 o K_1)(x, C) := \int_Y K_2(y, C) K_1(x, dy),\ x \in X,\ C \in \Sigma_Z$$

This integral is well-defined. For each fixed $C \in \Sigma_Z$, the map $y \longmapsto K_2(y, C)$ is Σ_Y-measurable. Hence it is integrable against the probability measure $K_1(x, dy)$ for every $x \in X$. Moreover, $(K_2 o K_1)(x, C) \in [0,1]$ since $K_2(y, C) \in [0,1]$ for all y. Thus composition is well-defined on objects and morphisms.

(ii) **Identity laws:** By proposition 4.4, we have $M o I_X = M = I_Y o M$

(iii) **Associative law:** By Proposition 4.3, for any three GRMs $M_1 = (X, Y, K_1)$, $M_2 = (Y, Z, K_2)$ and $M_3 = (Z, W, K_3)$, we have

$$M_3 o (M_2 o M_1) = (M_3 o M_2) o M_1$$

Thus the GRMs form a category, namely the Kleisli category of the Giry monad.■

Let $\mathrm{G}(Y)$ denote the set of probability measures on (Y, Σ_Y) with the σ-algebra generated by the evaluation maps $P \mapsto P(E)$. The assignment $Y \mapsto \mathrm{G}(Y)$ is the *Giry monad* on measurable spaces, with unit $\eta_Y(y) = \delta_y$ and multiplication $\mu_Y(\Pi)(E) = \int P(E)\, \Pi(dP)$ [6, 7]. On finite sets the same construction gives the finite distribution monad. Monads of this kind capture computations with effects ([11], [14]).

Proposition 4.5 *(GRMs are Kleisli arrows).* A function $K : X \times \Sigma_Y \to [0,1]$ is a Markov kernel if and only if $x \mapsto K(x, \cdot)$ is a measurable map $X \to \mathrm{G}(Y)$. Under this correspondence, Kleisli composition $g \circ_K f = \mu \circ \mathrm{G}(g) \circ f$ is the kernel composition (2.2), and the Kleisli identity η_X is the identity machine $I_X = (X, X, \delta)$ with kernel $\delta(x, A) = \mathbf{1}_A(x)$

Proof Measurability into $\mathrm{G}(Y)$ with the evaluation σ-algebra means that every $x \mapsto K(x, E)$ is measurable, which is condition (ii) of Definition 2.1; condition (i) says that each value is a probability measure. For arrows $f : X \to \mathrm{G}(Z)$ and $g : Z \to \mathrm{G}(Y)$ we have $(\mu \circ \mathrm{G}(g) \circ f)(x)(E) = \int g(z)(E)\, f(x)(dz)$, which is (2.2); and $\mu(\eta(P)) = P$ and $\mu(\mathrm{G}(\eta)(P)) = P$ give the unit laws. ∎

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Thus a GRM $M=(X,Y,K)$ is precisely an arrow $X\to Y$ in the Kleisli category of the Giry monad, called the category of stochastic maps or $\mathbf{Stoch}$ in [12]. A deterministic function f is lifted to this category as $\eta\circ f$, i.e. as the machine M_f.

4.3 Parallel composition and the monoidal structure

Definition 4.5 (Parallel Composition of Machines)

Let $M_1=(X_1,Y_1,K_1)$ and $M_2=(X_2,Y_2,K_2)$ be GRMs. Their *parallel composition* is the machine $M_1\otimes M_2=(X_1\times X_2,Y_1\times Y_2,K_1\otimes K_2)$, which runs the two machines side by side with independent randomness. Its kernel is the product kernel

$$(K_1\otimes K_2)\big((x_1,x_2),E\big)=\int\quad\int \mathbf{1}_E(y_1,y_2)\;K_1(x_1,dy_1)\;K_2(x_2,dy_2),\qquad (4.6)$$

so that $(K_1\otimes K_2)\big((x_1,x_2),E_1\times E_2\big)=K_1(x_1,E_1)\;K_2(x_2,E_2)$ on rectangles. By Tonelli's theorem this is a kernel on the product σ-algebra. For finite spaces the matrix of $M_1\otimes M_2$ is the Kronecker product $\mathrm{K}_1\otimes_{\mathrm{kr}}\mathrm{K}_2$.

Proposition 4.6 *(Interchange law).* For composable GRMs,
$(N_1\circ M_1)\otimes(N_2\circ M_2)=(N_1\otimes N_2)\circ(M_1\otimes M_2)$.

Proof. Evaluate both kernels on a rectangle $E_1\times E_2$ at (x_1,x_2). The left side equals $\int K_{N_1}(z_1,E_1)K_1(x_1,dz_1)\cdot\int K_{N_2}(z_2,E_2)K_2(x_2,dz_2)$. By Fubini's theorem the right side equals the same double integral. Rectangles form a π-system generating the product σ-algebra, so two probability measures that agree on rectangles coincide. ∎

Theorem 4.3 *(Symmetric monoidal structure).* With $\otimes$ on objects the Cartesian product, unit object the one-point space $\mathbf{1}=\{*\}$, and associator, unitors and symmetry given by the deterministic machines $M_\alpha,M_\lambda,M_\rho,M_\beta$ of the corresponding set-theoretic bijections, the category of measurable spaces and GRMs is a symmetric monoidal category.

Proof Bifunctoriality of $\otimes$ is Proposition 4.6 together with $I_{X_1}\otimes I_{X_2}=I_{X_1\times X_2}$. The assignment $f\mapsto M_f$ is a functor from measurable spaces and measurable functions to GRMs that preserves products, so the coherence isomorphisms of the Cartesian monoidal structure on measurable spaces are sent to isomorphisms satisfying the pentagon, triangle and

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

hexagon identities. Naturality of the associator and symmetry with respect to arbitrary GRMs follows by evaluating kernels on rectangles as in the previous proof. ∎

Proposition 4.7 *(Entropy is additive on independent machines).* For Finite Randomized Machines and inputs x_1, x_2, $H((M_1 \otimes M_2)(x_1, x_2)) = H(M_1(x_1)) + H(M_2(x_2))$.

Proof The output law is the product of $K_1(x_1, \cdot)$ and $K_2(x_2, \cdot)$, and entropy of a product measure is the sum of the entropies [19]. ∎

For a countable family of GRMs $M_i = (X_i, Y_i, K_i)$ the infinite product $\bigotimes_{i \geq 1} M_i = (\prod X_i, \prod Y_i, \bigotimes_{i \geq 1} K_i)$ exists, with kernel the infinite product of the kernels, by the existence theorem for product measures [1].

The category also has *copy* and *discard* maps. The copy machine $\Delta_X = (X, X \times X, K_\Delta)$ is the deterministic machine of $x \mapsto (x, x)$, and the discard machine $!_X = (X, \mathbf{1}, K_!)$ is the unique GRM into the one-point space. With these maps the category of GRMs between standard Borel spaces is a *Markov category* [12], so the general theory of conditional independence, sufficiency and Bayesian inversion developed there is available for GRMs. Our next result is an elementary characterization of determinism in these terms.

Theorem 4.4 *(Determinism is the copy rule).* Let Y be standard Borel. A GRM $M = (X, Y, K)$ is deterministic if and only if it commutes with copying:

$$(M \otimes M) \circ \Delta_X = \Delta_Y \circ M. \tag{4.7}$$

Proof. If $M = M_f$ is deterministic, both sides send x to $\delta_{(f(x), f(x))}$. Conversely, assume (4.7) and evaluate both sides at x on the set $E \times E$. The left side is $K(x, E)^2$, since the two copies are run independently, while the right side is $K(x, E)$, since both coordinates equal the same draw. Hence $K(x, E)^2 = K(x, E)$, so $K(x, E) \in \{0, 1\}$, and M is deterministic by Proposition 4.4. ∎

Intuitively, a machine that is run twice on a copied input gives two independent draws, whereas copying its *output* gives two identical ones; the two agree only when there is nothing random to copy.

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

5. Applications

5.1 Communication Theory (Noisy communication channel)

(i) Let X be an input alphabet and Y an output alphabet. Any discrete memoryless channel is the Finite Randomized Machine $M=(X,Y,K)$ with $K(x,y) := \mathbb{P}(Y=y|X=x)$. The binary symmetric channel (BSC) with crossover probability $\varepsilon\in[0,1/2)$ has $X=Y=\{0,1\}$ and $K(x,y)=(1-\varepsilon)\mathbf{1}_{y=x}+\varepsilon\mathbf{1}_{y\neq x}$.

(ii)The additive white Gaussian noise (AWGN) channel, $X=Y=\mathbb{R}$ with $K(x,\cdot)=\mathrm{N}(x,\sigma^2)$, is a General Randomized Machine (finite kernel replaced by a Markov kernel with Gaussian density).

5.2 Randomized Computation

Randomized algorithm: Let X be a finite set of problem instances and Y a finite set of outputs. The algorithm induces $K(x,y) := \mathbb{P}(\text{output } y|\text{input } x)$ ([16]). For a Monte Carlo algorithm with one-sided error, there exists a correct output $y^*(x)$ with $K\big(x,y^*(x)\big)\geq 1-\varepsilon$. For a Las Vegas algorithm, $K\big(x,y^*(x)\big)=1$, reducing M to a deterministic machine (Dirac kernel). The complexity class **BPP** is precisely the class of decision problems solvable by a Randomized Machine with $K\big(x,y^*(x)\big)\geq 2/3$ for all inputs .

Probabilistic Turing machine and automata:

(i) A Probabilistic Turing Machine ([10]) at state q reading symbol s samples the next configuration $(q',s',d)\in Q\times\Gamma\times\{L,R\}$. Setting $X=Q\times\Gamma$ and $Y=Q\times\Gamma\times\{L,R\}$, the one-step transition is a Finite Randomized Machine $M=(X,Y,K)$ with $K((q,s),(q',s',d))=\mathbb{P}(\text{transition to } (q',s',d)|(q,s))$. The row-sum condition $\sum_{(q',s',d)} K((q,s),(q',s',d))=1$ is the normalization of the transition distribution. The probabilistic Turing machine is thus a feedback network of this randomized machine (Theorem 4.1)

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

5.3 Stochastic Processes and Control

Markov chain: *A time-homogeneous Markov chain ([13]) on* $S=\{s_1,\ldots,s_N\}$ *with transition matrix* $P=(P_{ij})$ *is the Finite Randomized Machine* $M=(S,S,K)$ *with* $K(s_i,s_j) := P_{ij}$. *By Theorem 4.1, running* M *in feedback generates exactly this chain; the Markov property is not imposed but derived.*

Markov decision process: An MDP $(S,A,P_{\text{env}},r,\gamma)$ with $\gamma\in[0,1)$ ([9]) involves two randomized machines in sequential composition:

(i) The policy machine $M_\pi=(S,A,K_\pi)$ with $K_\pi(s,a)=\pi(a|s)$, and

(ii) The environment machine $M_{\text{env}}=(S\times A,S,K_{\text{env}})$ with $K_{\text{env}}((s,a),s')=P_{\text{env}}(s'|s,a)$.

Their composition $(K_{\text{env}}\circ K_\pi)(s,s')=\sum_a K_\pi(s,a)\,K_{\text{env}}((s,a),s')$ recovers the standard policy-induced transition matrix ([9]). The feedback network of this composed machine is the policy-induced state process $(S_t)_{t\geq 0}$.

5.4 Game Theory (Mixed strategies and game dynamics)

In a two-player finite game ([15]) , Player I's mixed strategy σ_1 over action set A_1 is the Finite Randomized Machine $M_1=(\{\text{play}\},\mathrm{A}_1,K_1)$ with $K_1(\text{play},a_j) := \sigma_1(a_j)$. In extensive-form games, a behavioral strategy at position x is $M=(\mathrm{P},\mathcal{M},K)$ where P is the set of positions and $\mathcal{M}$ the set of legal moves. Independent play is the parallel composition $M_1\otimes M_2$ with kernel $\sigma_1\otimes\sigma_2$, and the expected payoff

$u_i(\sigma_1,\sigma_2)=\sum_{a,b}\sigma_1(a)\,\sigma_2(b)\,u_i(a,b)$ is the mean of a deterministic payoff machine applied to its output. For sequential games such as chess let S be the set of positions, $\mathcal{M}$ the moves, $\tau: S\times\mathcal{M}\to S$ the deterministic rule applying a move, and $M_W=(S,\mathcal{M},K_W)$, $M_B=(S,\mathcal{M},K_B)$ the behavioral strategies. The composition of the two players' machines $M_{\mathrm{B}}\circ M_{\mathrm{W}}$ models a full move by both players; its feedback network is a Markov chain on game positions (Theorem 4.1), from which stationary distributions over terminal positions can be computed.

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

5.5 Machine Learning

Probabilistic classifier: A classifier maps feature vector $x \in X := X_{\text{feat}}$ to a distribution over class labels $Y = \{c_1, \ldots, c_k\}$ ([8]). With finite feature space this is a randomized machine $M = (X_{\text{feat}}, Y, K)$ with $K(x, c_j) := \mathbb{P}(c_j | x)$. Two standard instances are:

(i) *Logistic regression/ Neural network with softmax output* : $K(x, c_j) = \frac{e^{\mathbf{w}_j^\top x}}{\sum_{l=1}^{k} e^{\mathbf{w}_l^\top x}}$.

(ii) Autoregressive language model ([5]): Here $X := V^{\leq L}$ is the set of token-sequence prefixes, $Y := V$ is the vocabulary; $K(x, y)$ is the next-token probability given context x. Thus an autoregressive language model is the machine $M := M_{\text{LM}} = (V^{\leq L}, V, K)$ with context windows of length at most L as inputs and the vocabulary V as outputs. Sampling is the feedback of M_{LM} composed with the deterministic machine M_{app} that appends the new token and keeps the last L tokens. The sequence of context windows is therefore a time-homogeneous Markov chain on the finite set $V^{\leq L}$ with kernel $\sum_{w \in V:\, M_{\text{app}}(x, w) = x'} K(x, w)$. Thus Sequential composition of language model machines corresponds to autoregressive text generation.

5.6 Cognitive Science (Stimulus-response model)

In a stimulus–response model ([3]) a subject facing stimulus $s \in X$ *emits response* $r \in Y$ *with probability* $K_n(s, r)$ *on trial* n, *so each trial is a finite randomized machine* $M_n = (X, Y, K_n)$. *Learning acts on kernels; in a linear-operator rule of Bush–Mosteller type, a reinforced response* r_n *to stimulus* s *gives*

$$K_{n+1}(s, \cdot) = (1 - \theta)\, K_n(s, \cdot) + \theta\, \delta_{r_n}, \qquad 0 < \theta < 1$$

The reduced machine $\widetilde{M}$ (Theorem 3.4) collapses stimuli that produce the same response distribution, giving a canonical "effective stimulus space" from behavioral data alone.

International Journal of Engineering, Science and Mathematics
Vol. 15 Issue 11, Nov 2026,
ISSN: 2320-0294 Impact Factor: 8.215
Journal Homepage: http://www.ijesm.co.in, Email: ijesmj@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories
Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

5.7 Decision Theory (Decision under uncertainty and conditional generative model)

(i) With action set $X=A$ *and outcome set* $Y=O$*, define* $K(a,o) := \mathbb{P}(O=o|A=a)$ *. Expected utility ([4])* $E[u(M(a))]=\sum_o K(a,o)\,u(o)$ *is the composition of the machine's output with a deterministic utility machine* $M_u=(O,\mathbb{R},K_u)$ *where* $K_u(o,\cdot)=\delta_{u(o)}$*. The Bayesian optimal action* $a^*=argmax_{a\in A}(K_u\circ K)$ *is the optimal input to the composed machine* $M_u\circ M$*.*

(ii) A conditional generative model (VAE,GAN, or language model) ([8]) is a General Randomized Machine $M=(C,D,K)$ where C is a context space and $K(c,\cdot)=\mathbb{P}(D|C=c)$ is the generative distribution.

References

[1]O. Kallenberg, *Foundations of Modern Probability*, 2nd ed., Springer, New York, 2002.

[2]E. Çinlar, *Probability and Stochastics*, Graduate Texts in Mathematics 261, Springer, New York, 2011.

[3]R. R. Bush and F. Mosteller, Stochastic Models for Learning, Wiley, New York, 1955.

[4] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer, New

York, 1985.

[5]T. B. Brown et al., Language models are few-shot learners, *Adv. Neural Inf. Process. Syst.* **33** (2020), 1877–1901

[6]M. Giry, A categorical approach to probability theory, in *Categorical Aspects of Topology and Analysis*, Lecture Notes in Mathematics 915, Springer, Berlin, 1982, pp. 68–85.

[7]P. Panangaden, *Labelled Markov Processes*, Imperial College Press, London, 2009.

[8]C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.

[9]M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley, Hoboken, NJ, 2014.

[10] M. Sipser, *Introduction to the Theory of Computation*, 2nd ed., Thomson Course Technology, Boston, MA, 2006.

[11] S. Awodey, *Category Theory*, 2nd ed., Oxford University Press, Oxford, 2010.

[12] T. Fritz, A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics, *Adv. Math.* 370 (2020) 107239.

[13] J. R. Norris, *Markov Chains*, Cambridge University Press, Cambridge, 1997.

[14] E. Moggi, Notions of computation and monads, *Inform. and Comput.* 93(1) (1991) 55–92.

[15] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*, MIT Press, Cambridge, MA, 1994.

[16] R. Motwani and P. Raghavan, *Randomized Algorithms*, Cambridge University Press, Cambridge, 1995.

[17] H. Bauer, *Measure and Integration Theory*, de Gruyter Studies in Mathematics 26, Walter de Gruyter, Berlin, 2001.

[18] D. Pollard, *A User's Guide to Measure Theoretic Probability*, Cambridge University Press, Cambridge, 2002.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley-Interscience, Hoboken, NJ, 2006.