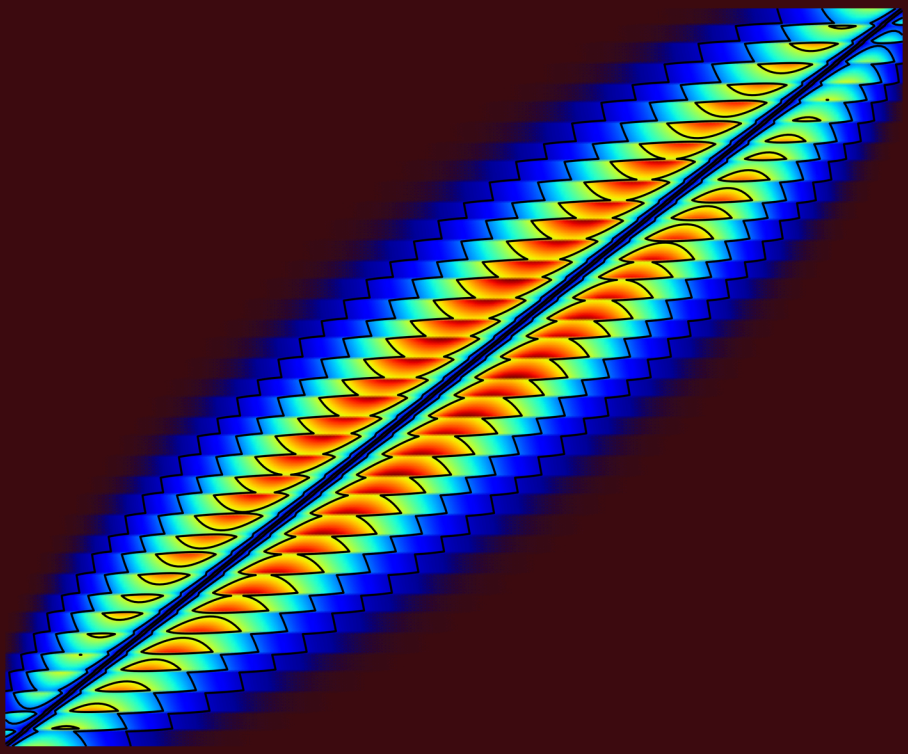


Exact binomial decision regret.

Stochastic Filtering for Partially Observed Diffusions

Filtering Equations,
Density Methods,
and Particle Approximations

Erik Lindevall

ODIN PRESS

Stochastic Filtering for Partially Observed Diffusions

Copyright © 2026 by **Erik Lindevall**.
All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the prior written permission of the publisher, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law.

Paperback version, First edition: 2026.
ISBN: forthcoming (KDP allocation pending)
Trim Size: 6×9 inch. Paper color: White.

Algorithm copyright. All algorithms reproduced in this volume, in the chapter bodies and their pseudocode boxes, are the copyrighted intellectual property of **Erik Lindevall**. No reproduction, redistribution, adaptation or incorporation into any derivative work is permitted without the prior written permission of the copyright holder.

Published by Odin Press.
`https://odinpress.org`
This book was typeset in LaTeX by Odin Press from a manuscript submitted by the author.

9 8 7 6 5 4 3 2 1 0

Preface

A measurement leaves a hidden system partly unknown. What it changes is the distribution of the states that remain possible, and stochastic filtering studies that distribution when the hidden system is a diffusion and the observations arrive through a separate noisy channel. The conditional mean is one summary of it; event probabilities, credible regions and decisions draw on the rest of the law, and this book keeps the whole law in view.

The first three chapters build the target. Chapter 1 fixes the observer's information as a filtration and constructs regular conditional laws relative to it. Chapter 2 develops the signal, its generator and the prediction operator, with exact Gaussian transitions for linear signals in any dimension. Chapter 3 adds Bayes updates at sampled times and turns the recursion into two exact algorithms, the forward algorithm on a finite state set and the sampled Kalman filter. Chapter 4 then proves the Kalman–Bucy filter for a continuous record by Gaussian orthogonality and Itô's formula alone, and studies the Riccati flow as a dynamical system: its comparison theorem, the algebraic Riccati equation, and convergence to the stabilising solution under detectability and stabilisability.

Chapters 5 and 6 treat the general model. A reference probability makes the record a Brownian motion independent of the signal, the Kallianpur–Striebel formula writes the filter as a ratio, and the unnormalised filter solves the linear Zakai equation, which has exactly one solution for regular coefficients. The robust representation of Clark and Davis expresses the filter as a continuous functional of the observed path. Innovations and Itô's quotient rule then lead to the Kushner–Stratonovich equation. Chapter 7 asks when finitely many statistics carry the conditional law: it proves the Wonham equation for finite-state chains and the Beneš filter, a nonlinear diffusion whose filter is a reweighted Kalman–Bucy filter, and it places the cubic sensor beyond any smooth finite-dimensional recursion. Chapter 8 studies how a filter forgets its initial law, with exponential rates for the

Kalman–Bucy filter and Birkhoff's contraction coefficient in the Hilbert metric for positive kernels.

The last four chapters approximate. Chapter 9 writes the conditional density as the solution of a stochastic partial differential equation, converts it through the robust form into a parabolic equation with random coefficients, and computes it by the splitting-up scheme, which alternates a Fokker–Planck step on a grid with an exact Bayes factor for each observation increment. Chapter 10 projects the density onto finite families; there the theorem of Brigo, Hanzon and Le Gland identifies the Hellinger projection filter on an exponential family with the Stratonovich assumed-density filter, and the extended Kalman filter appears as a linearised Gaussian closure. Chapter 11 builds particle filters from importance weights and the Feynman–Kac recursion, with adaptive and systematic resampling, auxiliary and Rao–Blackwellised filters, an unbiased estimator of the evidence and Girsanov-weighted particles for a continuous record. Chapter 12 proves that these filters converge, with a constant that stays bounded in time when the signal mixes, and carries the approximation error through to the risk of a decision.

The reader is assumed to know measure-theoretic probability, Itô calculus, linear algebra and elementary partial differential equations. Results are proved where a graduate reader can follow the proof in a few pages, and cited with their hypotheses where the full argument belongs to a monograph. The texts of Bain and Crisan, Liptser and Shiryaev, Kallianpur and Jazwinski are good companions throughout.

The figures are drawn from exact formulas and chosen to make a mechanism visible: a posterior whose two modes leave its mean in a valley, a Riccati variance approaching its floor from both sides, weights that concentrate while particle labels disappear. Every chapter states at least one algorithm as a procedure with its inputs, its steps and its output, so that each method can be implemented as it is read.

The chapters are meant to be read in order through Chapter 8. After that, the density chapters and the particle chapters form two connected routes into approximation, and Chapter 12 brings them together.

Erik Lindevall

Contents

Preface vii

List of Figures xiii

List of Tables xvii

1 Conditional Distributions and Partial Observation 1

1.1 Signals, observations and available information 2
1.2 Conditional expectation as an information projection . 5
1.3 Regular conditional laws and test functions 9
1.4 Filtering, prediction and smoothing horizons 15
1.5 Bayes risk and the cost of a point estimate 18
1.6 A complete conditional law before an algorithm 22

2 Signal Models and Prediction 33

2.1 A well-posed hidden diffusion 34
2.2 Transition kernels and prediction with fixed information 37
2.3 The generator as an Itô identity 42
2.4 Weak forward evolution and a separate norm topology 45
2.5 Exact OU kernels and changing density shape 50
2.6 An exact prediction algorithm and its boundaries . . . 58

3 Bayesian Updates and Discrete Observations 67

3.1 Measurement kernels and predictive evidence 68
3.2 Reweighting, support and finite posterior moments . . 71
3.3 Prediction between sampled observations 76
3.4 The quadratic sensor and a posterior with two modes . 81
3.5 An analytic diffusion and mixture update 87

3.6 Endpoint sensors, observation increments and the forward algorithm 91

4 Linear Gaussian Filtering 101
4.1 Linear observations and Gaussian conditional geometry 102
4.2 Conditioning on an observation path 105
4.3 Matrix Riccati flow and covariance order 111
4.4 An exact scalar Ornstein–Uhlenbeck benchmark 119
4.5 Density geometry and the limits of Gaussian closure . 124
4.6 Exact recursions for sampled observations 129

5 Reference Measures and Unnormalised Filters 135
5.1 A reference law that separates signal and observations 136
5.2 Likelihoods require a true martingale 139
5.3 Conditional Bayes formulas and unnormalised mass . . 143
5.4 Projecting a product decomposition onto observations . 146
5.5 The weak Zakai equation 150
5.6 A static signal makes the likelihood visible 161

6 Innovations and the Nonlinear Filtering Equation 169
6.1 Prediction removed from the observation 170
6.2 Martingale representation and equality of filtrations . . 173
6.3 The Ito quotient and the normalised equation 177
6.4 What the normalised equation preserves and reveals . . 183
6.5 Analytic Gaussian checks of the stochastic coefficients . 187
6.6 From exact identities to a usable filtering procedure . . 195

7 Finite-State Filters and Finite-Dimensional Closure 203
7.1 A finite-state comparison to diffusion filtering 204
7.2 The Wonham probability equation 208
7.3 Why the probability simplex is invariant 212
7.4 Exact finite-state recursions and their observation models 218
7.5 What exact finite-dimensional closure requires 224
7.6 Limits of closure and the route to approximation . . . 232

8 Stability, Forgetting and Model Error 237
8.1 Comparing conditional laws on the same observations . 238
8.2 Exact forgetting in the scalar Gaussian filter 242

8.3 Positive kernels and an explicit contraction certificate . 251
8.4 Observability, nondegeneracy and the hypotheses of general results 256
8.5 Persistent model error and misleading concentration . 262
8.6 What observation diagnostics can establish 267

9 Conditional Densities and Their Equations 273
9.1 When the conditional law admits a density 274
9.2 The adjoint generator and density transport 278
9.3 Zakai as a density-valued stochastic equation 280
9.4 Normalisation and the density innovation term 288
9.5 Reflecting, killed and truncated state domains 292
9.6 Exact density fields and the splitting-up scheme 296

10 Density Approximations and Projection Filters 307
10.1 Targets, embeddings and approximation geometry . . . 308
10.2 Galerkin coefficients, mass and signed truncations . . . 311
10.3 Hellinger projection and the stochastic calculus 317
10.4 Assumed densities and forward entropy projection . . . 322
10.5 Residuals and the limits of local accuracy 331
10.6 An exact quartic posterior and its Gaussian projection 334

11 Particle Approximations and Resampling 341
11.1 A conditional law represented by weighted points . . . 342
11.2 Importance weights and what their concentration measures 344
11.3 Feynman–Kac weights and the sampled filtering recursion 348
11.4 Transition proposals and observation-adapted proposals 351
11.5 Resampling preserves an empirical mean and adds randomness 359
11.6 A complete sampled bootstrap filter and its diversity limits 364

12 Approximation Guarantees and Decisions under Partial Observation 375
12.1 Fixed-grid particle estimates and finite-time bounds . . 376
12.2 Test functions, finite loss classes and exact error laws . 386

12.3 From continuous observations to a frozen-likelihood grid . . . 389
12.4 Separating discretisation, projection and particle errors . . . 395
12.5 Posterior actions, risk perturbations and decision margins . . . 399
12.6 A capstone guarantee and the limits of conditional inference . . . 404

References **409**

Index **421**

About the Author **425**

Chapter 1

Conditional Distributions and Partial Observation

A measuring instrument seldom reports the quantity we want. It reports a noisy, accumulated or partial function of it: a position blurred by sensor noise, a magnitude stripped of its sign, a voltage that depends on a hidden state through a nonlinear response. This book asks what such a record allows us to conclude about the hidden state, and how to compute that conclusion as the record grows in continuous time. The answer is a probability law, the conditional distribution of the state given the observations. A single estimate is extracted from it afterwards, once a loss has been chosen, and different losses extract different estimates from the same law.

This first chapter builds the objects that the rest of the book computes. It fixes the observer's information as the filtration generated by the observed path and shows that every admissible estimate is a functional of that path. It constructs conditional expectations and regular conditional laws relative to this information, proves the Bayes formula, and separates filtering, prediction and smoothing by two time indices. It then turns a conditional law into decisions through conditional risks. A scalar Gaussian mixture carries all of these objects in closed form, including a conditional mean that sits between two posterior modes, and leads to an exact finite decision algorithm.

The later chapters supply the dynamics and the equations: signal models and prediction in Chapter 2, Bayes updates at sampled times in Chapter 3, the Kalman–Bucy, Zakai, Kushner–Stratonovich and Wonham equations

for continuous records in Chapters 4 to 7, the stability of filters in Chapter 8, and their approximation by densities, projections and particles in Chapters 9 to 12.

1.1 Signals, observations and available information

The hidden signal $X = (X_t)_{t\geq 0}$ and the observation $Y = (Y_t)_{t\geq 0}$ are two random processes on one probability space. The model describes both; an estimate may use only the second. This is the setting of the classical monographs of Liptser and Shiryaev [1], Kallianpur [2], Jazwinski [3] and Bain and Crisan [4], which the reader will find useful alongside every chapter of this book, while the measure theory of conditioning used below is developed in Chapter 6 of Kallenberg [5]. For a state in $\mathbb{R}^d$ and a record in $\mathbb{R}^m$, the following assumptions give the continuous-time setting.

Assumption 1.1 (A probability model for partial observation)**.** Work on a complete filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ carrying continuous adapted processes X and Y with values in $\mathbb{R}^d$ and $\mathbb{R}^m$. Let $h : \mathbb{R}^d \to \mathbb{R}^m$ be Borel measurable, let R be a constant positive definite m by m matrix, and let V be an m-dimensional Brownian motion relative to the model filtration. Assume $Y_0 = 0$ and $Y_t = \int_0^t h(X_s)\,ds + R^{1/2}V_t$, that V is independent of the initial signal and of the signal's driving noise, and that $E\int_0^T \|h(X_s)\|^2\,ds < \infty$ for every finite T. Chapter 2 constructs signals with these properties.

Here $R^{1/2}$ is the positive definite square root of R. The integrability condition gives the observation drift its meaning: by the Cauchy–Schwarz inequality in time, $E\int_0^T \|h(X_s)\|\,ds$ is at most $\sqrt{T}$ times the square root of $E\int_0^T \|h(X_s)\|^2\,ds$. A continuous adapted process is progressively measurable, so $h(X_s)$ is jointly measurable in (s, ω) on every finite interval. The observed random object is the integrated process Y; the familiar notation $dY_t = h(X_t)\,dt + R^{1/2}dV_t$ abbreviates the integral equation and involves no derivative of a Brownian path.

Two cases of the assumption recur in this chapter. With $h \equiv 0$ the record is independent of the signal, and Section 1.4 uses it to separate a change of target from a change of information. With a static scalar signal, $h(x) = x$ and $R = r^2$, the value $Y_t/t = X + rV_t/t$ is a single noisy measurement of X with noise variance r^2/t; Section 1.6 studies this one-time measurement in full.

The model filtration may contain X_0, the signal noise and the observation noise up to time t. That is what allows the joint model to be written down, and it makes these quantities available to the analysis of the model, while the observer receives only Y. Defining V on the same space as Y leaves its realized path hidden: recovering V from Y would require subtracting the observation drift, which depends on the hidden signal. The information used by an estimate is therefore a filtration in its own right, as in the continuous-time formulation of [6] (Chapter 7 and Section 7.2).

Definition 1.2 (Raw, completed and right-limit observation filtrations)**.** The raw observation filtration is $\mathcal{Y}_t^0 := \sigma(Y_s : 0 \le s \le t)$. Let $\mathcal{N}^P$ contain every subset of every P-null event of $\mathcal{F}$, and put $\mathcal{Y}_t := \mathcal{Y}_t^0 \vee \mathcal{N}^P$. This completed observation filtration is the conditioning information of this book unless another choice is stated. Its right-limit augmentation is $\mathcal{Y}_{t+} := \bigcap_{u>t} \mathcal{Y}_u$; equality of $\mathcal{Y}_t$ and $\mathcal{Y}_{t+}$ is a property of a model, proved where it is used.

An observation filtration is increasing, since an earlier record is part of every later one. Completion and the right limit are different operations: completion adds the subsets of probability-zero events, while the right limit intersects the information available at all strictly later times. For the models of Chapters 5 and 6, right-continuity of $\mathcal{Y}_t$ is proved through an equivalent reference law under which Y is a Brownian motion with covariance R; Chapters 5 and 6 give the argument. The present chapter conditions at fixed times, where only completion enters. Because the model filtration is complete, $\mathcal{Y}_t \subseteq \mathcal{F}_t$, and a $\mathcal{Y}_t$-measurable real random variable is called an admissible estimate at time t.

For continuous paths the raw history is generated by countably many values: each Y_s with $s < t$ is the limit of Y_q along rational $q < t$ approaching s, and the endpoint Y_t is added. Conditioning on a whole path is therefore conditioning on countably many coordinates. The effect of completion is elementary and useful.

Proposition 1.3 (Completion preserves events modulo null sets)**.** For fixed t, every $A \in \mathcal{Y}_t$ differs from some $A_0 \in \mathcal{Y}_t^0$ by a P-null set. Hence the two indicators define the same element of $L^p(P)$ for every $1 \le p < \infty$.

Proof. Let $\mathcal{A}$ be the class of events $A \in \mathcal{F}$ for which some $A_0 \in \mathcal{Y}_t^0$ satisfies $P(A \triangle A_0) = 0$. Taking complements preserves the symmetric difference. If A_n has raw representative $A_{0,n}$, the symmetric difference of $\bigcup_n A_n$ and

Chapter 3

Bayesian Updates and Discrete Observations

A tracking sensor reports the position of a diffusing particle once every few seconds, and each report carries an independent error. Each report raises two questions: what is the conditional law of the particle now, and how probable was the report under the model? One repeated operation answers both. Between reports the conditional law moves with the transition law of the signal. At a report it is multiplied by the likelihood of the reading and divided by the predictive density of that reading. This chapter makes that operation exact.

Chapter 1 defined the conditional law as the object of filtering, and Chapter 2 supplied the transition law of the hidden diffusion. Here the observations arrive at isolated times. We first establish Bayes' formula for a measurement kernel, with the predictive evidence as its denominator, and study what reweighting preserves: null sets, ratios of masses, moments, and distances between laws. With Markov prediction, the update gives an exact recursion for the sampled filter and an unnormalised form whose mass is the likelihood of the data. Two representations make the recursion a finite computation: a finite state space, where it is the forward algorithm, and the linear Gaussian signal, where it is the sampled Kalman filter. A quadratic sensor and a Gaussian mixture give exact nonlinear posteriors, and the first has a conditional mean lying between two modes. The final section separates a sampled endpoint reading from an increment of a continuous record, which is the observation model of the chapters that follow.

3.1 Measurement kernels and predictive evidence

Let X denote a hidden state in a standard Borel space E, with prior probability μ, and let Z be a reading in a standard Borel space F. A measurement kernel specifies the distribution of Z at each possible state, before any reading is supplied.

Assumption 3.1 (Endpoint observation model)**.** Let ν be a sigma-finite measure on F. Given $X = x$, the observation Z has a jointly measurable, nonnegative density $g(z, x)$ satisfying $\int_F g(z, x)\,\nu(dz) = 1$ for every $x \in E$. In a repeated endpoint model, observations are conditionally independent given the sampled state sequence, and the conditional distribution of each observation depends only on its corresponding state.

For one measurement the joint probability is determined on rectangles by

$$\mathbb{P}(X \in A, Z \in C) = \int_A \int_C g(z, x)\,\nu(dz)\,\mu(dx).$$

The state and observation measures have separate roles. The prior may be any probability on E, with or without a density, while the observation density is integrated against ν. Counting measure gives a categorical sensor; Lebesgue measure gives a real-valued reading. This is the state-space convention of Doucet and Johansen, Section 2.1, equations (1)–(7) [15], written with kernels so that singular state laws are included. Cappé, Moulines and Rydén develop hidden Markov models in the same kernel generality [16].

Definition 3.2 (Discrete measurement likelihood)**.** For a fixed reading $z \in F$, the discrete measurement likelihood is $\ell_z(x) := g(z, x)$, regarded as a function of state. Its prior integral $c_\mu(z) := \mu(\ell_z) = \int_E g(z, x)\,\mu(dx)$ is the predictive evidence at z.

The subscript on c_μ records the prior used to predict the reading; in the recursion below the same letter carries the step index. In a continuous observation space, $g(z, x)$ and $c_\mu(z)$ are density values. Probabilities come from integrating them over observation sets.

Proposition 3.3 (Evidence is an observation density)**.** Under the endpoint observation model, c_μ is measurable and $\int_F c_\mu(z)\,\nu(dz) = 1$. The random evidence $c_\mu(Z)$ is finite and strictly positive almost surely.

Proof. Apply Tonelli's theorem to the nonnegative jointly measurable function g. Measurability of parameter integrals gives measurability of c_μ, and interchange of integration gives

$$\int_F c_\mu(z)\,\nu(dz) = \int_E \left(\int_F g(z,x)\,\nu(dz) \right) \mu(dx) = 1.$$

The joint model, integrated over all states, also gives $\mathbb{P}(Z \in C) = \int_C c_\mu(z)\,\nu(dz)$. Thus the set on which $c_\mu = 0$ has observation probability zero. Since its integral is finite, c_μ is infinite only on a ν-null set, which also has observation probability zero. □

The proposition concerns a reading drawn from the model: the exceptional set $\{c_\mu = 0\} \cup \{c_\mu = \infty\}$ has observation probability zero. A reading supplied from outside the model, through a sensor fault or a transcription error, can fall in it, and so the denominator is evaluated at the data before any division.

Theorem 3.4 (Bayes formula for a kernel)**.** For each z satisfying $0 < c_\mu(z) < \infty$, define $\mu^z(dx) := \ell_z(x)\mu(dx)/c_\mu(z)$. This is a probability measure. Extend it by any measurable probability assignment on the exceptional observation set. The resulting kernel is a regular conditional law of X given Z. At an admissible z, its expectation of a measurable real function φ is $\mu^z(\varphi) = \mu(\varphi\ell_z)/c_\mu(z)$ whenever $\mu(|\varphi|\ell_z) < \infty$.

Proof. Fix an admissible z. Integration of a nonnegative function against μ defines a measure, and its mass after division by $c_\mu(z)$ is one. For each measurable $A \subset E$, the map $z \mapsto \mu^z(A)$ is measurable on the admissible set by measurability of numerator and denominator. A constant probability assignment, for example μ itself, supplies a measurable extension; any other measurable probability assignment serves equally well.

Write $D := \{z : 0 < c_\mu(z) < \infty\}$ and $\Gamma(A; z) := \int_A g(z,x)\,\mu(dx)$. For a measurable observation set C, the proposed kernel satisfies

$$\int_C \mu^z(A)\,\mathbb{P}_Z(dz) = \int_{C \cap D} \Gamma(A; z)\,\nu(dz).$$

On $\{c_\mu = 0\}$ the kernel $\Gamma(A; \cdot)$ vanishes, because it is bounded by $\Gamma(E; \cdot) = c_\mu$, and $\{c_\mu = \infty\}$ is ν-null. We can therefore replace $C \cap D$ by C. The result is $\mathbb{P}(X \in A, Z \in C)$, the defining disintegration identity. Applying this identity to nonnegative simple functions and taking increasing limits proves the expectation formula for nonnegative tests; applying it separately to positive and negative parts proves the integrable case. □

What does an observation allow us to conclude about a hidden diffusion?

The answer is a conditional distribution. This book follows that distribution from conditional expectation and Bayesian updating to the equations of nonlinear filtering, then to the approximations needed when exact closure fails.

Kalman–Bucy filtering provides an exact Gaussian benchmark. Reference measures, the Zakai equation and innovations lead to the general filtering equations. Later chapters develop stability, conditional densities, projection methods and particle approximations, keeping the target law and the assumptions of each construction explicit.

Analytic contours, worked examples, proofs and full algorithm descriptions connect the theory to concrete calculations. The final chapter separates approximation errors and examines how errors in a posterior affect a decision.

For graduate students, researchers and mathematically trained practitioners with a background in probability, Itô calculus, linear algebra and elementary differential equations.



ISBN 13 978-983-43-5247-9
Paperback
9 789834 352479