1

Wiener Processes

The Wiener process is one of the easiest models for random accumulation of degradation over time. It is based on the assumption of an additive accumulation of degradation with linear wear intensity. Regarding every degradation increment as an additive superposition of a large number of small effects, we can assume the degradation process to be normally distributed.

1.1. Gaussian distribution

Since the Wiener process is based on normally distributed increments, we recall here some basic properties of the Gaussian (normal) distribution.

DEFINITION 1.1.– *A random variable X is said to be (univariate) Gaussian distributed, or normally distributed, with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$ if its distribution admits the following probability density function (p.d.f.) with respect to Lebesgue measure:*

$$\forall x \in \mathbb{R}, \quad f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.$$

By extension, a constant random variable $X = \mu$ with $\mu \in \mathbb{R}$ is said to be Gaussian distributed with mean μ and variance $\sigma^2 = 0$. In any case, we write $X \sim \mathcal{N}(\mu, \sigma^2)$.

The specific case $\mathcal{N}(0,1)$ where $\mu = 0$ and $\sigma = 1$ is called the standard Gaussian (normal) distribution.

DEFINITION 1.2.– *An n-dimensional random vector $X = (X_1, \ldots, X_n)^T$ is said to be (multivariate) Gaussian distributed (or normally distributed or*

Degradation Processes in Reliability, First Edition. Waltraud Kahle, Sophie Mercier and Christian Paroissin.

a Gaussian vector) if any linear combination of its components is a normally distributed random variable:

$$\forall a_1, \ldots, a_n \in \mathbb{R}, \quad a_1 X_1 + \cdots + a_n X_n$$

is (univariate) normally distributed.

A direct consequence of the definition is that Gaussian distributions are stable under marginalization: Any subvector of a Gaussian vector still is a Gaussian vector. In particular, all the components of a Gaussian vector are Gaussian distributed random variables. Also, any random vector that is obtained through affine transformation of a Gaussian vector is a Gaussian vector. Furthermore, using the Fourier transform, it can be checked that the distribution of a Gaussian vector X only depends on its mean vector μ and covariance matrix Σ. The corresponding distribution is denoted by $\mathcal{N}_n(\mu, \Sigma)$, where n stands for the dimension of X. We now get the following result, where the symbol ^{T} means transposition.

PROPOSITION 1.3.– Let $X = (X_1, \ldots, X_n)^T$ be a Gaussian vector with $X \sim \mathcal{N}_n(\mu, \Sigma)$. Let A be a $p \times n$ matrix and B a $p \times 1$ column vector. Then, $AX + B$ is a Gaussian vector with distribution $\mathcal{N}_p(A\,\mu + B, A\,\Sigma\,A^T)$.

We can also check that gathering independent Gaussian random variables $X_1, \ldots, X_n$ provides a Gaussian vector $(X_1, \ldots, X_n)$. As a specific case, gathering n independent and identically distributed (i.i.d.) standard Gaussian random variables provides a standard Gaussian vector with distribution $\mathcal{N}_n(0_n, I_n)$, where 0_n refers to an n-dimensional vector and I_n to the n-dimensional identity matrix.

The previous results provide a way to construct a random vector with a specific Gaussian distribution.

COROLLARY 1.4.– Let Y be a n-dimensional standard Gaussian vector ($Y \sim \mathcal{N}_n(0, I_n)$). Let Σ be a symmetric and positive semi-definite $n \times n$ matrix and μ be an n-dimensional vector. Let A be a matrix such that $\Sigma = AA^T$. Then, the random vector $X = A\,Y + \mu$ is a Gaussian vector with mean μ and covariance matrix Σ.

Assuming further that Σ is non-singular (and hence symmetric positive definite), it is easy to derive from the previous corollary that a Gaussian vector $X \sim \mathcal{N}_n(\mu, \Sigma)$ admits the following p.d.f. with respect to Lebesgue measure:

$$\forall x \in \mathbb{R}^n, \quad f_X(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left\{ -\frac{1}{2}(x-\mu)^T\, \Sigma^{-1} (x-\mu) \right\},$$

where $x = (x_1, \ldots, x_n)^T$, Σ^{-1} is the inverse matrix of Σ and $\det \Sigma = |\Sigma|$ is its determinant.

Whether Σ is non-singular or not, an important property of a Gaussian vector is that its components are independent random variables as soon as the covariance matrix Σ is diagonal. As mentioned in the Introduction, the Gaussian distribution is also infinitely divisible.

All the processes that we consider in the rest of this chapter are Gaussian processes, which we now define.

DEFINITION 1.5.– *A time-continuous stochastic process $X = (X_t)_{t\geq 0}$ is said to be a Gaussian process if, for any $n \in \mathbb{N}^*$ and for any $(t_1, \ldots, t_n) \in \mathbb{R}_+^n$, the random vector $(X_{t_1}, \ldots, X_{t_n})$ is a Gaussian vector.*

1.2. Brownian motion

We here introduce and provide basic properties of the well-known Brownian motion, which is a specific Wiener process and from where general Wiener processes will next be constructed.

DEFINITION 1.6.– *$B = (B_t)_{t\geq 0}$ is a Brownian motion if, and only if,*

1) *$B_0 = 0$ a.s.;*

2) *B has stationary increments: for any $t, s, h > 0$, $B_t - B_s \overset{d}{=} B_{t+h} - B_{s+h}$;*

3) *B has independent increments: for any $n \geq 1$ and for any $t_1 < \cdots < t_n$, the random variables B_{t_1}, $B_{t_2} - B_{t_1}$, …, $B_{t_n} - B_{t_{n-1}}$ are independent;*

4) *for any $t > 0$, $B_t \sim \mathcal{N}(0, t)$.*

Figure 1.1 shows some simulations of Brownian motion paths.

The Brownian motion has a number of remarkable properties:

1) the Brownian motion is a Gaussian process. In particular, $(B_{t_1}, \ldots B_{t_n}) \sim \mathcal{N}_n(\mu, \Sigma)$ with mean $\mu = 0_n$ and covariance matrix $\Sigma = (\min\{t_i, t_j\})_{i,j\in\{1,\ldots,n\}^2}$;

2) the Brownian motion is a Lévy process and it is the only Lévy process with a.s. continuous sample paths;

3) the sample paths are almost surely not differentiable anywhere and have unbounded variation;

4) reflection property. If the path of a Brownian motion reaches a value a at time $t = s$, then the subsequent path after time s is identically distributed as the horizontally reflected path, starting from point a at time s;

5) self-similarity. For $0 < t_1 < t_2 < \cdots < t_n$ and $T > 0$

$$(\sqrt{T}B_{t_1}, \sqrt{T}B_{t_2}, \ldots, \sqrt{T}B_{t_n}) \stackrel{d}{=} (B_{Tt_1}, B_{Tt_2}, \ldots, B_{Tt_n}).$$

For an extensive description of Brownian motion and its properties, the reader could refer to [MÖR 10], for instance.

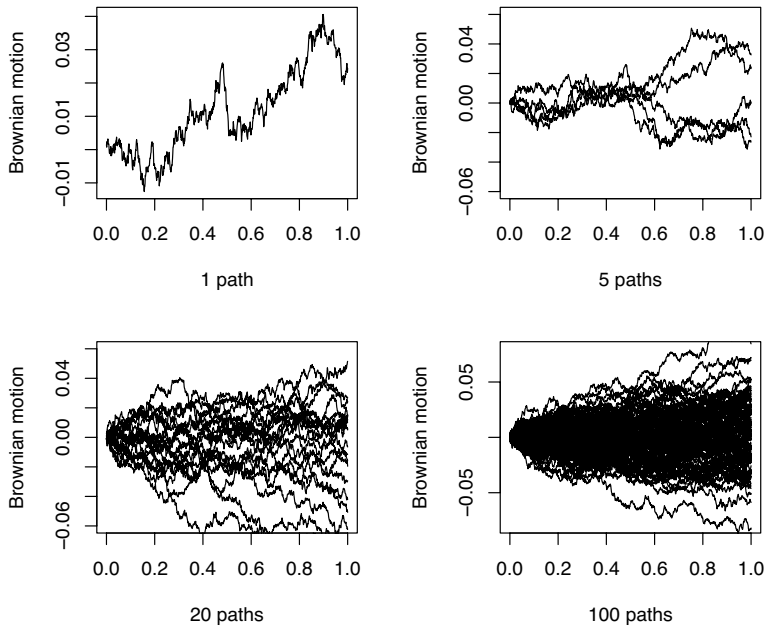

Figure 1.1. *Simulations of Brownian motion path*

1.3. Constructions and simulation methods

In this section, we provide three constructions of the Brownian motion, leading to different simulation methods.

1.3.1. *Random walk approximation*

Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with zero mean and unit variance. We consider the random walk $(R_n)_{n\geq 0}$ defined by:

$$R_0 = 0 \quad \text{and} \quad R_n = X_1 + \cdots + X_n, \quad n \geq 1.$$

Then, by the central limit theorem (see theorem 4.4, page 120 in [PET 95], for instance),

$$\sup_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{R_n - \mathbb{E}R_n}{\sqrt{\mathrm{Var}(R_n)}} \leq x \right) - \Phi(x) \right| \xrightarrow[n\to\infty]{} 0\,,$$

where Φ is the cumulative distribution function (c.d.f.) of the standard Gaussian distribution (i.e. with zero mean and unit variance). Consider the process with continuous sample paths on $[0,1]$: [MIK 98]

$$S_n(t) = \begin{cases} R_i/\sqrt{n} & \text{for } t = \frac{i}{n} \text{ with } i \in \{1,\ldots,n\} \\ \text{linearly interpolated elsewhere.} \end{cases}$$

Let us consider, for the moment, the restriction of the process S_n at points i/n and let us assume that the distribution of $X_1, X_2, \ldots$ is the standard Gaussian distribution. Then, the following properties hold:

– $\mathbb{P}(S_n(0) = 0) = 1$;

– $(S_n(t))_{t\geq 0}$ has independent increments in the following sense: for any $(i_1,\ldots,i_m) \in \mathbb{N}^m$ with $m < n$ and $0 \leq i_1 < \ldots < i_m = n$,

$$S_n(i_2/n) - S_n(i_1/n), \ldots,\ S_n(i_m/n) - S_n(i_{m-1}/n)$$

are independent;

– for all $0 \leq i \leq n$, $S_n(i/n) \sim \mathcal{N}(0, i/n)$.

Hence, $S_n(t)$ and B_t for $t \in \{0, 1/n, 2/n, \ldots, 1\}$ share a lot of identical properties.

By the *functional central limit theorem*, also called *Donsker's invariance principle* [BIL 99, theorem 8.2, p. 90], the process $(S_n(t))_{t\geq 0}$ converges uniformly on $C[0,1]$, the space of continuous functions on the unit interval, to

the Brownian motion on $[0,1]$. For all possible choices of $t_i \in [0,1]$ and $x_i \in \mathbb{R}$, for $i \in \{1,\ldots,m\}$ with $m \geq 1$,

$$\mathbb{P}(S_n(t_1) \leq x_1, \ldots, S_n(t_m) \leq x_m) \xrightarrow[n\to\infty]{} \mathbb{P}(B_{t_1} \leq x_1, \ldots, B_{t_m} \leq x_m).$$

From the self-similarity property of the Brownian motion, we also know how to obtain an approximation of the sample paths of the Brownian motion on any interval $[0,T]$: simulate one path of $(S_n)_{n\geq 0}$ on $[0,1]$, then scale the time interval by the factor T and the sample path by the factor $\sqrt{T}$. Thus, a first simple tool for simulating Brownian sample paths is given by the following algorithm.

ALGORITHM 1.7.– Simulation of a Brownian motion based on random walk approximation.

For $0 = t_0 < t_1 < \cdots < t_k = T$, generate standard normally distributed random variables $X_1, \ldots, X_k$ and recursively define

$$\begin{aligned} B_{t_1} &= \sqrt{t_1} X_1 \\ B_{t_2} &= B_{t_1} + \sqrt{t_2 - t_1} X_2 = \sqrt{t_1} X_1 + \sqrt{t_2 - t_1} X_2 \\ &\vdots \\ B_{t_k} &= \sum_{i=1}^{k} \sqrt{t_i - t_{i-1}} X_i \end{aligned}$$

The paths shown in Figure 1.1 were simulated by this method.

1.3.2. *Brownian bridge sampling*

The Brownian bridge construction is a way to build a path of the Brownian motion by successively adding finer scale details: to simulate a Brownian motion on $[0,T]$, we first simulate a realization of Z_T. Next, we simulate a realization of $Z_{T/2}$ given Z_T and $Z_0 = 0$. Next, a realization of $Z_{T/4}$ given Z_0 and $Z_{T/2}$ is simulated, together with a realization of $Z_{3T/4}$ given $Z_{T/2}$ and Z_T, and so on. At the n-th step, the interval $[0,T]$ is divided into 2^n intervals of the form:

$$I_{n,k} = \left[\frac{k-1}{2^n}T, \frac{k}{2^n}T\right] = [t_{n,k-1}, t_{n,k}], \quad k \in \{1,\ldots,2^n\}.$$

These intervals are chosen in such a way that going from level n to level $n+1$ divides each interval of the n-th level exactly in half.

The Brownian bridge construction builds an approximated Brownian motion path B_t for $t \in [0,T]$ iteratively and level by level. At level n, the path $B_{n,t}$ is constructed through a piecewise linear approximation. This piecewise linear path is determined by its values $B_{n,t_{n,k}}$ at the endpoints $t_{n,k}$, with $k \in \{0,\dots,2^n\}$. The construction ensures that these $B_{n,t_{n,k}}$ have the same joint distribution as the corresponding $B_{t_{n,k}}$ of an exact Brownian motion path. Going from stage n to stage $n+1$, one keeps the values $B_{n,t_{n,k}}$ and introduces a new value at the center of each interval $I_{n,k}$. To simulate these new values, we have to calculate the conditional distribution of B_t at the midpoint given the two boundary values. This is done through the following proposition. Note that a Brownian motion in an interval with given boundary values is called *Brownian bridge*.

PROPOSITION 1.8.– Let $0 < s < u < t$. Then, given $B_s = x$ and $B_t = z$, the random variable B_u is Gaussian distributed with

$$\mathbb{E}(B_u | B_s = x, B_t = z) = x + \frac{z-x}{t-s}(u-s)$$

and

$$\mathrm{Var}(B_u | B_s = x, B_t = z) = \frac{(u-s)(t-u)}{t-s}$$

for all $x, z \in \mathbb{R}$.

PROOF.– Let $f_{u|s,t}(\cdot|x,z)$ be the conditional p.d.f. of B_u given $B_s = x$ and $B_t = z$, with

$$\forall y \in \mathbb{R}, \quad f_{u|s,t}(y|x,z) = \frac{f_{s,u,t}(x,y,z)}{f_{s,t}(x,z)},$$

where $f_{s,u,t}$ and $f_{s,t}$ stand for the p.d.f. of (B_s, B_u, B_t) and (B_s, B_t), respectively.

Because of the independent increments of the Brownian motion, we have:

$$f_{s,u,t}(x,y,z) = f_s(x) f_{u-s}(y-x) f_{t-u}(z-y)$$

and

$$f_{s,t}(x,y,z) = f_s(x) f_{t-s}(z-x),$$

where, for all $t > 0$, f_t stands for the p.d.f. of the Gaussian distribution with zero mean and variance t, with

$$\forall x \in \mathbb{R}, \quad f_t(x) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right).$$

Therefore, the conditional p.d.f. $f_{u|s,t}(y|x,z)$ can be calculated as follows:

$$\begin{aligned} f_{u|s,t}(y|x,z) =& \frac{f_{u-s}(y-x) \cdot f_{t-u}(z-y)}{f_{t-s}(z-x)} \\ =& \frac{1}{\sqrt{2\pi}} \sqrt{\frac{t-s}{(u-s)(t-u)}} \exp\left(-\frac{A}{2}\right) \end{aligned}$$

with

$$\begin{aligned} A =& \frac{(y-x)^2}{u-s} + \frac{(z-y)^2}{t-u} - \frac{(z-x)^2}{t-s} \\ =& \frac{1}{(t-s)(u-s)(t-u)} \left((y-x)(t-u) - (z-y)(u-s)\right)^2 \\ =& \frac{(t-s)}{(u-s)(t-u)} \left(y - x - \frac{z-x}{t-s}(u-s)\right)^2. \end{aligned}$$

It is now easy to recognize that $f_{u|s,t}(\cdot|x,z)$ is the p.d.f. of the Gaussian distribution with the mean and variance given in the proposition. □

For simulation purposes through the Brownian bridge, we need the conditional distribution of B_u given $B_s = x$ and $B_t = z$ at the midpoint, that is for $u = (t+s)/2$. From the previous result, it is easy to see that it is a Gaussian distribution with expectation $(x+z)/2$ and variance $(t-s)/4$. This leads to the following algorithm.

ALGORITHM 1.9.– [Simulation of a Brownian motion based on Brownian bridge sampling] Generate standard normally distributed random variables $X_1, X_2, \ldots$

– Step 0. Set $t_{0,0} = 0$, $t_{0,1} = T$ and $B_{t_{0,0}} = 0$, $B_{t_{0,1}} = \sqrt{T} X_1$.

– Step 1. Leave $B_{t_{1,0}} = B_{t_{0,0}}$, $B_{t_{1,2}} = B_{t_{0,1}}$. Set $\Delta t = T$ and add

$$B_{t_{1,1}} = (B_{t_{1,0}} + B_{t_{1,2}})/2 + \frac{\sqrt{\Delta t}}{2} X_2.$$

– Step $n+1$. Leave all $B_{t_{n+1,2k}} = B_{t_{n,k}}$, for any $k \in \{0, \ldots, 2^n\}$. Set $\Delta t = T/2^n$ and add, for all $k \in \{1, \ldots, 2^n\}$ (i.e. for each interval), the value

$$B_{t_{n+1,2k-1}} = (B_{t_{n+1,2k-2}} + B_{t_{n+1,2k}})/2 + \frac{\sqrt{\Delta t}}{2} X_j,$$

where X_j is the next value in the sequence ahead.

The advantage of Brownian bridge sampling is that it is not necessary to give the distance between time points in advance. Further, by this method additional process values can be added to an already simulated sample path, if necessary. Figure 1.2 shows one simulated sample path for $n = 4, 5, 6$ (left) and $n = 10$ (right).

1.3.3. *Karhunen–Loève approximation*

In the previous algorithms, each simulated sample path is a piecewise linear function. In the theory of stochastic processes, the Karhunen–Loève theorem is a representation of a stochastic process as an infinite linear combination of orthogonal functions, analogous to a Fourier series representation of a function on a bounded interval. In contrast to a Fourier series where the coefficients are fixed numbers and the expansion basis consists of sinusoidal functions, the coefficients in the Karhunen–Loève theorem are random variables and the expansion basis depends on the process.

In the case of a centered stochastic process $(Y_t)_{t\in[a,b]}$, Y_t admits a decomposition

$$Y_t = \sum_{k=1}^{\infty} Z_k e_k(t),$$

where $(Z_k)_{k\geq 1}$ is a sequence of pairwise uncorrelated random variables, and where $(e_k)_{k\geq 1}$ is a sequence of continuous real-valued functions on $[a, b]$ that are pairwise orthogonal in $L^2([a, b])$. Moreover, if the process is Gaussian, then the random variables Z_k are Gaussian and independent.

The Karhunen–Loève expansion of a Brownian motion in $[0, T]$ is given by (see [ASH 90], for instance):

$$B_t = \sum_{k=1}^{\infty} Z_k e_k(t) \quad \text{with} \quad e_k(t) = \frac{2\sqrt{2T}}{(2k+1)\pi} \sin\left(\frac{(2k+1)\pi t}{2T}\right) \qquad [1.1]$$

where $(Z_k)_{k\geq 1}$ is a sequence of i.i.d. standard Gaussian random variables. To approximate a sample path of the Brownian motion, we can truncate the series in equation [1.1] up to the N-th term. Figure 1.3 shows some simulated sample paths for different N.

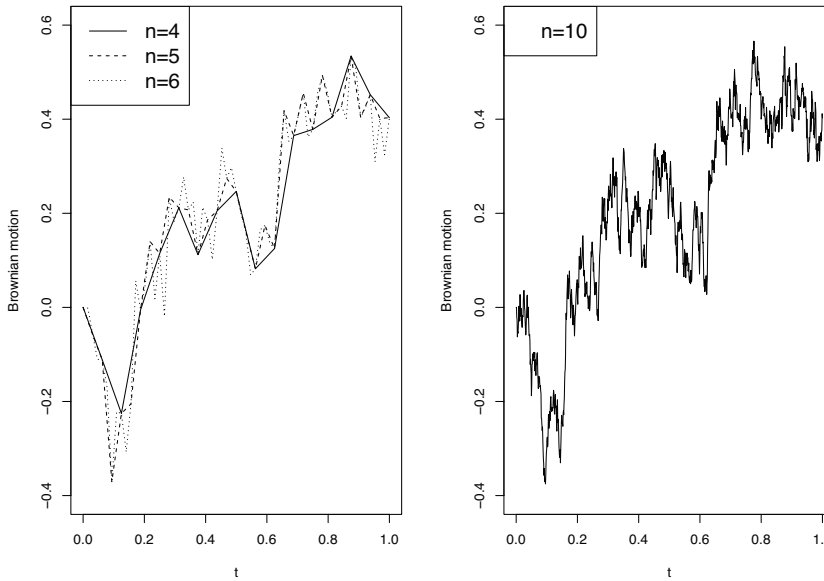

Figure 1.2. *Brownian bridge sampling*

REMARK 1.10.– All three methods of simulation give good results. The usage depends on the concrete situation. The random walk approximation is simple and fast, but the times for process values must be given in advance. In the Brownian bridge sampling, it is possible to add process values to an already simulated sample path, if necessary. By these two methods, the process values are simulated at discrete time points. This is the major difference to the Karhunen–Loève approximation where the sample path is not piecewise linear.

1.4. Wiener process (Brownian motion with drift)

A Wiener process is a Brownian motion with an additional linear drift function and an additional variance parameter:

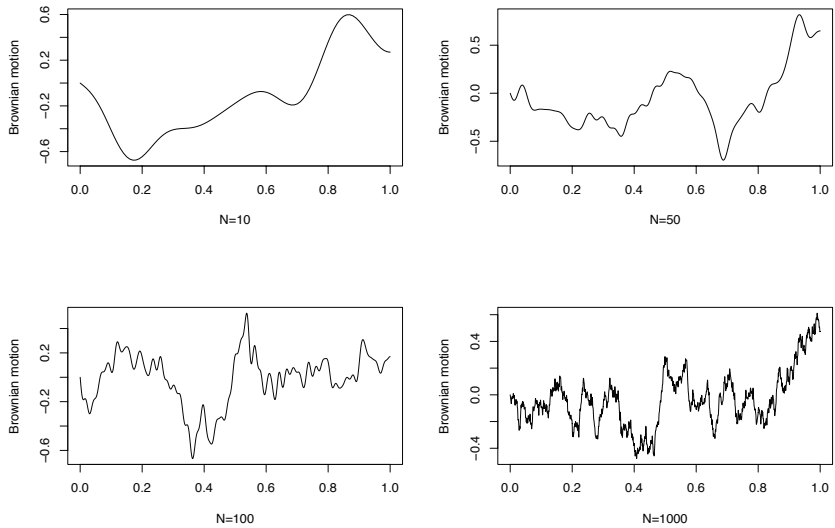

Figure 1.3. *Karhunen–Loève approximation for the Brownian motion*

DEFINITION 1.11.– *The stochastic process* $W = (W_t)_{t\geq 0}$ *is called* a Wiener process *if*

$$\forall t \geq 0, \quad W_t = \mu t + \sigma B_t, \tag{1.2}$$

where $\mu \in \mathbb{R}$, $\sigma > 0$ *and* $(B_t)_{t\geq 0}$ *is a Brownian motion:* μ *is a drift parameter (that we expect to be positive in our case) and* σ^2 *is a variance parameter (or volatility parameter).*

The Wiener process with drift μ has the same properties as the Brownian motion. In this way, the increments remain independent. The only difference is some change in the parameters of the Gaussian distributed increments: For all $s < t$, the increment $W_t - W_s$ is Gaussian distributed with

$$\mathbb{E}(W_t - W_s) = \mu(t-s) \quad \text{and} \quad \mathrm{Var}(W_t - W_s) = \sigma^2(t-s). \tag{1.3}$$

It is sometimes useful to make the model more flexible by introducing two additional parameters x_0 and t_0, with $x_0,\ t_0 \in \mathbb{R}$. Then the degradation measure $(X_t)_{t\geq 0}$ can be defined by the following model:

$$\forall t \geq t_0, \quad X_t = x_0 + W_{t-t_0} = x_0 + \mu(t-t_0) + \sigma B_{t-t_0}, \tag{1.4}$$

where $x_0 \in \mathbb{R}$ is a constant initial degradation, $t_0 \in \mathbb{R}$ is the time when the degradation accumulation begins and $(W_t)_{t\geq 0}$ is a Wiener process.

Model [1.4] differs from [1.2] in the time where the degradation begins to accumulate. It changes the initial time from 0 to t_0 and the initial state from 0 to x_0. The influence of the parameters x_0 and t_0 is illustrated in Figure 1.4. If $t_0 < 0$, then at time $t = 0$ (time when the use of the item begins), the state of the item is random. More specifically, it is Gaussian distributed with mean $x_0 + \mu|t_0|$ and variance $\sigma^2|t_0|$. Assuming h to be a critical level for the deterioration, we will see later on that it is not the boundary h itself, but the difference $h - x_0$ (degradation reserve) that plays an important role. When $t_0 < 0$, the random state at time $t = 0$ provides a random degradation reserve at the same time. An alternate possibility is to directly model h as a random boundary. On the other side, if $t_0 > 0$, the time t_0 can be seen as an "incubation time". It is also possible that we begin the observation of the process after it has been running for a while, and that we are interested in its future development. If there is a short period of nonlinear behavior at the beginning, as is plotted in Figure 1.4(b), and if the observation begins after that time, when the mean function has become linear, then the short nonlinearity period is absorbed by (probably false) estimates of x_0 and t_0, which are not of further interest. This makes the model more flexible.

For any $t \geq t_0$, the p.d.f. of X_t is then

$$\forall x \in \mathbb{R}, \quad f_{X_t}(x) = \varphi\left(\frac{x - x_0 - \mu(t - t_0)}{\sigma(t - t_0)}\right), \qquad [1.5]$$

where φ is the p.d.f. of the standard Gaussian distribution.

1.5. First passage time (time to failure)

We suppose that a failure of a product will occur if the degradation process arrives at a certain critical degradation level, which may be unknown. For a given constant boundary h, the lifetime τ_h of the product is then determined as the first instant at which the degradation process $(X_t)_{t\geq 0}$ exceeds the level h:

$$\tau_h = \inf\{t \geq t_0; X_t \geq h\}. \qquad [1.6]$$

In case $x_0 \geq h$, it is readily seen that $\tau_h = t_0$. We consequently only consider the case where $x_0 < h$ in the following.

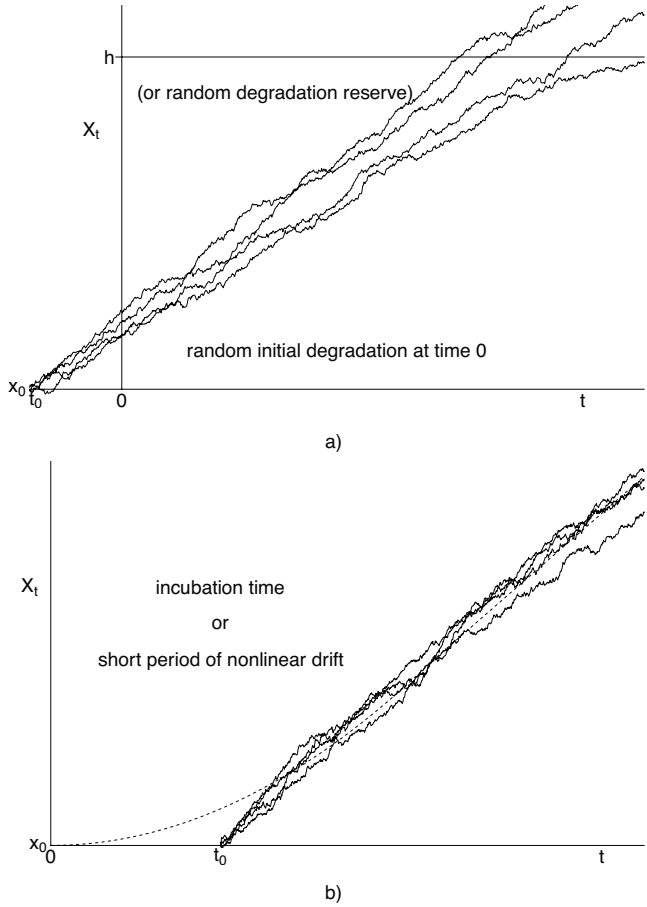

Figure 1.4. *Random initial degradation a) or incubation time b)*

Let $t > t_0$. Noticing that $X_t > x$ with $x > h$ implies that $\tau_h < t$, the formula of total probability implies that, for any $x > h$:

$$\mathbb{P}(X_t > x) = \mathbb{P}(X_t > x; \tau_h < t) = \int_{t_0}^{t} \mathbb{P}(X_t > x | \tau_h = s) \, \mathrm{d}F_{\tau_h}(s).$$

Based on the continuity of $(X_t)_{t\geq 0}$, we also have $X_{\tau_h} = h$. Using furthermore the independent and stationary increments property, we obtain

$$\begin{aligned}\mathbb{P}(X_t > x) &= \int_{t_0}^{t} \mathbb{P}(X_t > x | X_s = h)\, \mathrm{d}F_{\tau_h}(s) \\ &= \int_{t_0}^{t} \mathbb{P}(X_t - X_s > x - h)\, \mathrm{d}F_{\tau_h}(s) \\ &= \int_{t_0}^{t} \mathbb{P}(X_{t-s} > x - h)\, \mathrm{d}F_{\tau_h}(s). \end{aligned} \quad [1.7]$$

Equation [1.7] says that the Wiener process is above state $x > h$ at time t if it takes s units of time to reach level h for the first time and next, the increment during a time interval with length $t - s$ is above $x - h$ (see Figure 1.5).

Taking the derivative of [1.7] with respect to x, we get an integral equation for the p.d.f. of the first passage time:

$$f_{X_t}(x) = \int_{t_0}^{t} f_{\tau_h}(s) f_{X_{t-s}}(x - h)\, \mathrm{d}s, \quad [1.8]$$

where $f_{X_{t-s}}$ is the p.d.f. of the increment of the Wiener process during time $t - s$ and f_{τ_h} the p.d.f. of τ_h.

Let us first consider the solution of the integral equation [1.7] (or equivalently of [1.8]) for the special case $\mu = 0$, $x_0 = 0$ and $t_0 = 0$ (which corresponds to a Brownian motion).

PROPOSITION 1.12.– For $\mu = 0$, $x_0 = 0$ and $t_0 = 0$,

$$\forall t \geq 0, \quad F_{\tau_h}(t) = 2\left(1 - \Phi\left(\frac{h}{\sigma\sqrt{t}}\right)\right)$$

is the unique solution of equation [1.7]. The p.d.f. is then given by:

$$\forall t \geq 0, \quad f_{\tau_h}(t) = \frac{h}{\sqrt{2\pi\sigma^2 t^3}} \exp\left(-\frac{h^2}{2\sigma^2 t}\right). \quad [1.9]$$

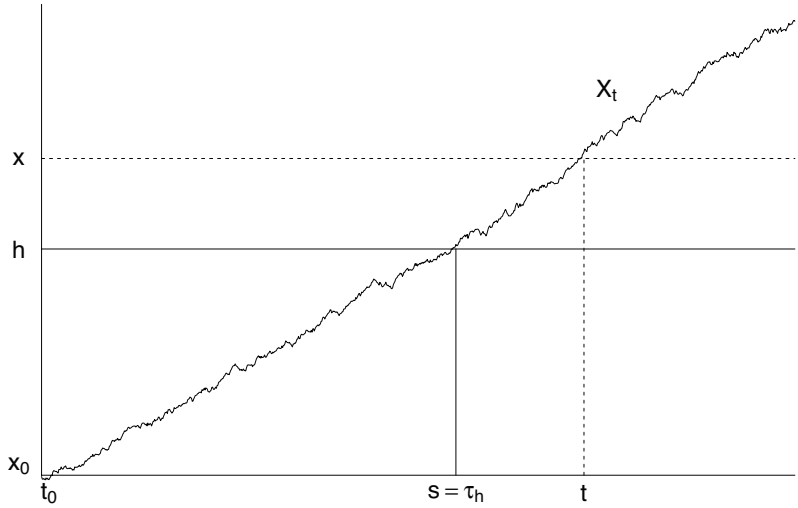

Figure 1.5. *A sample path of the Wiener process*

PROOF.– For these parameters, we have $X_t = \sigma B_t$ and, since $\frac{B_t}{\sqrt{t}}$ is standard Gaussian distributed,

$$\mathbb{P}(X_t > x) = \mathbb{P}(\sigma B_t > x) = \mathbb{P}\left(\frac{B_t}{\sqrt{t}} > \frac{x}{\sigma\sqrt{t}}\right) = 1 - \Phi\left(\frac{x}{\sigma\sqrt{t}}\right),$$

where Φ is the cumulative distribution function of the standard Gaussian distribution. Furthermore, for $x = h$, we have

$$\mathbb{P}(X_{t-s} > x - h) = F_{X_{t-s}}(0) = \frac{1}{2},$$

because X_{t-s} is Gaussian distributed with zero expectation. Now, we get, from equation [1.7],

$$\mathbb{P}(X_t > h) = \int_0^t \frac{1}{2}\,\mathrm{d}F_{\tau_h}(s) = \frac{1}{2}F_{\tau_h}(t)$$

and the c.d.f. of the first passage time is

$$F_{\tau_h}(t) = 2\,\mathbb{P}(X_t > h) = 2\left(1 - \Phi\left(\frac{h}{\sigma\sqrt{t}}\right)\right).$$

Finally, taking the derivative, the p.d.f. of τ_h is given by:

$$\forall t \geq 0, f_{\tau_h}(t) = \frac{\mathrm{d}F_{\tau_h}(t)}{\mathrm{d}t} = \frac{h}{\sigma\sqrt{t^3}}\Phi'\left(\frac{h}{\sigma\sqrt{t}}\right) = \frac{h}{\sqrt{2\pi\sigma^2 t^3}}\exp\left(-\frac{h^2}{2\sigma^2 t}\right).$$

□

From equation [1.9], we can observe that the first passage time of a Brownian motion is inverse Gaussian distributed.

DEFINITION 1.13.– *The inverse Gaussian distribution (also known as the Wald distribution) is a two-parameter family of continuous probability distributions with support on* $(0, \infty)$. *Its p.d.f. is given by:*

$$\forall t \geq 0, \quad f(t) = \sqrt{\frac{\beta}{2\pi t^3}}\exp\left(\frac{-\beta(t-\mu_1)^2}{2\mu_1^2 t}\right), \qquad [1.10]$$

where $\mu_1 > 0$ *is the mean and* $\beta > 0$ *is the shape parameter.*

It is much more complicated to develop the p.d.f. of the first passage time in the case of a general Wiener process with drift. However, the previous result on the Brownian motion gives us an idea of the general solution and we now show that the first passage time for a fixed level $h > x_0$ of a Wiener process with positive drift is still inverse Gaussian distributed.

THEOREM 1.14.– Remembering that the distribution of X_t is given by equation [1.5], then

$$\forall t \geq t_0, \quad f_{\tau_h}(t) = \frac{h - x_0}{\sqrt{2\pi\sigma^2(t-t_0)^3}}\exp\left(-\frac{(h - x_0 - \mu(t-t_0))^2}{2\sigma^2(t-t_0)}\right). \qquad [1.11]$$

PROOF.– First, we prove that [1.11] fulfills equation [1.8]. For this, we have to show that

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi\sigma^2(t-t_0)}}\exp\left(-\frac{(x - x_0 - \mu(t-t_0))^2}{2\sigma^2(t-t_0)}\right) \\
&= \int_{t_0}^{t} \frac{h - x_0}{\sqrt{2\pi\sigma^2(s-t_0)^3}}\exp\left(-\frac{(h - x_0 - \mu(s-t_0))^2}{2\sigma^2(s-t_0)}\right) \\
&\qquad \times \frac{1}{\sqrt{2\pi\sigma^2(t-s)}}\exp\left(-\frac{(x - h - \mu(t-s))^2}{2\sigma^2(t-s)}\right)\mathrm{d}s.
\end{aligned}$$

The right-hand side of this equation can be written as:

$$\frac{1}{\sqrt{2\pi\sigma^2(t-t_0)}}\exp\left(-\frac{(x-x_0-\mu(t-t_0))^2}{2\sigma^2(t-t_0)}\right)\times I,$$

where

$$I=\int_{t_0}^{t}\frac{1}{\sqrt{2\pi\sigma^2}}\frac{(h-x_0)(t-t_0)^{1/2}}{(t-s)^{1/2}(s-t_0)^{3/2}}\exp\left(-\frac{1}{2\sigma^2}\cdot A(s)\right)\,\mathrm{d}s,$$

and

$$\begin{aligned}A(s) &= \frac{(h-x_0)^2}{(s-t_0)}+\frac{(x-h)^2}{(t-s)}-\frac{(x-x_0)^2}{(t-t_0)}\\ &= \frac{((x-h)(s-t_0)-(h-x_0)(t-s))^2}{(s-t_0)(t-s)(t-t_0)}\\ &= \frac{((x-h)\frac{(s-t_0)}{(t-s)}-(h-x_0))^2}{\frac{(s-t_0)}{(t-s)}(t-t_0)}.\end{aligned}$$

Applying the substitution $u=\frac{s-t_0}{t-s}$ for $t_0<s<t$, that is $s=\frac{t_0+ut}{1+u}$ with $0<u<\infty$, we obtain:

$$I=\int_0^{\infty}\frac{1}{\sqrt{2\pi u^3}}\frac{(h-x_0)}{\sigma(t-t_0)^{1/2}}\exp\left(-\frac{(x-h)^2(u-\frac{(h-x_0)}{(x-h)})^2}{2\sigma^2(t-t_0)u}\right)\mathrm{d}u.$$

It is easy to see that this is an integral over the complete support of an inverse Gaussian distribution with $\beta=(h-x_0)^2/(\sigma^2(t-t_0))$ and $\mu_1=(h-x_0)/(x-h)$. Therefore, $I=1$ and [1.11] fulfills equation [1.8].

We now prove that the solution of [1.8] is unique. This can be shown by using the Laplace transforms of $t\mapsto f_{\tau_h}(t)$ and $t\mapsto f_{X_t}(x)$:

$$\mathcal{L}_{\tau_h}(s;x_0,h)=\int_{t_0}^{\infty}e^{-st}f_{\tau_h}(t)\,\mathrm{d}t,\quad \mathcal{L}_X(x;x_0,s)=\int_{t_0}^{\infty}e^{-st}f_{X_t}(x)\,\mathrm{d}t.$$

Note that $\mathcal{L}_{\tau_h}(s)$ is a moment generating function whereas $\mathcal{L}_{X_t}(s)$ is not, the integration being taken with respect to t. The integral in the right-side of equation [1.8] can be seen as a convolution product. As the Laplace transform

of a convolution product is the product of the Laplace transforms, we derive that

$$\mathcal{L}_{\tau_h}(s; x_0, h) = \frac{\mathcal{L}_X(x; x_0, s)}{\mathcal{L}_X(x; h, s)}.$$

There is hence one single possible Laplace transform of f_{τ_h} for f_{τ_h} to be a solution of [1.8]. The uniqueness of the solution of [1.8] hence follows from the one-to-one property of Laplace transform. See [CHH 89] for more details and a proof in a more general setting (positive, negative, and zero drift). □

We now make some remarks and give some properties of the p.d.f. given in equation [1.11]. More about the inverse Gaussian distribution can be found, for example, in [CHH 89] or in [JØR 82].

1) Setting

$$\begin{cases} \beta &= \frac{(h-x_0)^2}{\sigma^2} \\ \mu_1 &= \frac{h-x_0}{\mu} \end{cases} \iff \begin{cases} \sigma^2 &= \frac{(h-x_0)^2}{\beta} \\ \mu &= \frac{h-x_0}{\mu_1} \end{cases}, \qquad [1.12]$$

we can see that equations [1.11] and [1.10] are equivalent for $t_0 = 0$ so that, indeed, τ_h is inverse Gaussian distributed with parameter (β, μ_1) in the sense of definition 1.13. $t_0 \in \mathbb{R}$ is an additional location parameter and we get the three-parameter inverse Gaussian distribution (see [PAD 79] for instance).

2) The c.d.f. of τ_h is given by, for any $t \geq t_0$,

$$F_{\tau_h}(t) = \mathbb{P}(\tau_h \leq t) = \Phi\left(\frac{\mu(t-t_0) - h + x_0}{\sigma\sqrt{t-t_0}}\right) + \exp\left(\frac{2\mu(h-x_0)}{\sigma^2}\right)\Phi\left(-\frac{\mu(t-t_0) + h - x_0}{\sigma\sqrt{t-t_0}}\right).$$

This can be checked by taking its derivative that leads to equation [1.11].

3) The characteristic function $C_{\tilde{\tau}_h}(z) = \mathbb{E}(e^{iz\tilde{\tau}_h})$, where $\tilde{\tau}_h = \tau_h - t_0$, is given by

$$C_{\tilde{\tau}_h}(z) = \exp\left(\frac{\beta}{\mu_1}\left(1 - \left(1 - \frac{2i\mu_1^2 z}{\beta}\right)^{1/2}\right)\right),$$

where (β, μ_1) is provided by [1.12]. Taking the k-th derivative of $C_{\tilde{\tau}_h}(z)$ and evaluating it at $z = 0$, we get (see, e.g. [CHH 89]):

$$\mathbb{E}[\tilde{\tau}_h^k] = \mu_1^k \sum_{u=0}^{k-1} \frac{(k-1+u)!}{u!((k-1-u)!} \left(\frac{\mu_1}{2\beta}\right)^u. \qquad [1.13]$$

It follows that the expectation and the variance of τ_h are given by:

$$\mathbb{E}(\tau_h) = t_0 + \mu_1 = t_0 + \frac{h - x_0}{\mu} \text{ and } \mathrm{Var}(\tau_h) = \frac{\mu_1^3}{\beta} = \frac{(h-x_0)\sigma^2}{\mu^3}.$$

Note that for a Brownian motion without drift, that is $\mu = 0$, the expectation of τ_h does not exist.

4) The failure rate $r(t) = f(t)/(1 - F(t))$ is rather complicated, but it is not difficult to compute it for given parameters. The behavior of the failure rate was analyzed in detail in [CHH 89]. It is increasing for $t < t^*$ and decreasing for $t > t^*$, where t^* is the solution of

$$r(t) = \frac{\beta}{2\mu_1^2} + \frac{3}{2t} - \frac{\beta}{2t^2}.$$

Furthermore, it can be shown that

$$\lim_{t\to\infty} r(t) = \frac{\beta}{2\mu_1^2}.$$

At first look, it might seem surprising that the inverse Gaussian distribution does not have an increasing failure rate. The reason becomes clearer if we think in terms of the corresponding parameters of the Wiener process. In [CHH 89, p. 153], some failure rates are plotted for $\mu_1 = 1$ and different β. It can be seen that the mode t^* of $r(t)$ is small for β small. Since $\beta = (h-x_0)^2/\sigma^2$, the mode of $r(t)$ is hence small either if the distance between the initial value x_0 and the boundary h is small, or if the variance σ^2 of the process is very large. In both cases, the stochastic process either reaches the boundary level after a short time or it first goes down and needs some time to come back. The decreasing property of the failure rate after a short time can hence be justified in terms of the physical behavior: excluding the items that are reaching the boundary after a short time, the failure rate of the surviving items will be relatively small. In Figure 1.6, some failure rates together with their c.d.f. for $h - x_0 = 5$, $\mu = 1$ and different σ are plotted. It can be seen that for small and moderate

variances σ^2, the failure rate is increasing, it is not monotonous any more when σ^2 is large.

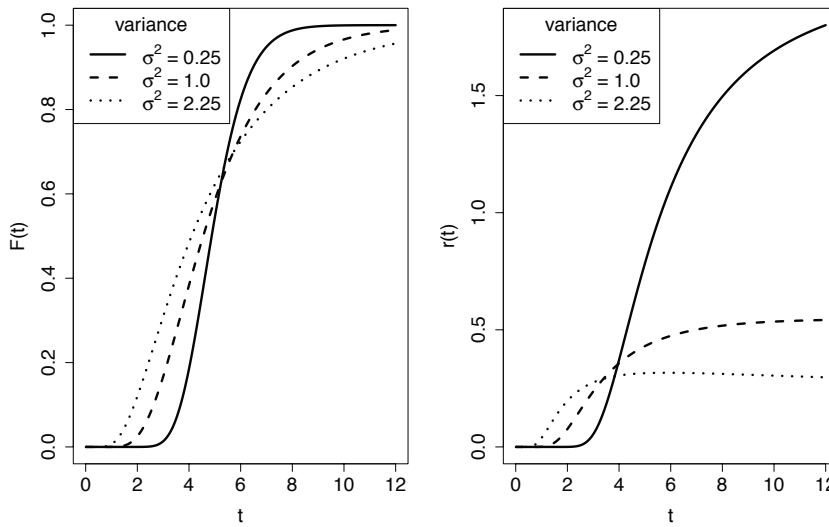

Figure 1.6. *C.d.f. and failure rate of some times-to-failure distributions*

5) Let $\tilde{\tau}_h = \tau_h - t_0$. In [CHH 89], it was shown that for the mean residual life $\mathbb{E}(\tilde{\tau}_h - T|\tilde{\tau}_h > T)$ of an item with age T, the following properties hold:

$$\lim_{T\to 0} \frac{\mathbb{E}(\tilde{\tau}_h - T|\tilde{\tau}_h > T)}{\mathbb{E}(\tilde{\tau}_h)} = 1 \text{ and}$$

$$\lim_{T\to \infty} \frac{\mathbb{E}(\tilde{\tau}_h - T|\tilde{\tau}_h > T)}{\mathbb{E}(\tilde{\tau}_h)} = \frac{2\mu_1}{\beta} = \frac{2\sigma^2}{\mu(h - x_0)}.$$

When σ^2 is large, the mean residual lifetime can hence exceed the mean initial lifetime of the system (obtained when $T \to 0$). This apparently surprising property can be explained in a similar way as for the failure rate.

1.6. Statistical inference

It was shown that for $x_0 < h$, the lifetime τ_h follows an inverse Gaussian distribution with p.d.f. given in equation [1.11]. The parameters

of this distribution are expressed with respect to those of the degradation process using [1.12] so that estimating the parameters of the inverse Gaussian distribution is mostly equivalent to estimating the parameters of the Wiener process (details further). The degradation level h may also be an unknown parameter, which may hence have to be estimated too. To estimate the parameters of this model, several observations schemes are possible:

1) Only increments of the degradation process are observed. In this case, the likelihood function of the observation is a product of Gaussian densities. It is possible to estimate the parameters of the degradation process μ, σ^2, x_0 and t_0, but it is impossible to estimate the degradation level h. Note that, neither the same delay between successive observation times, nor identical observation times, are required. Parameter estimations can be obtained even if the observation times are all distinct.

2) Only failure times are observed. In this case, we get the classical likelihood function from a sample of inverse Gaussian distributed random variables. From this likelihood function, we can estimate the parameters μ, σ^2 and t_0. The degradation reserve $h - x_0$ must be given in advance.

3) The sample includes both process increments and failure times. Assume first that the sample consists of N failure times and of $\sum_{i=1}^{n} m_i$ increments corresponding to different degradation processes from those on which the failure times are observed. Then the likelihood function is the product of the likelihood functions mentioned above. In the case where failure times and degradation measures correspond to the same process, the likelihood function is more complicated to be computed. In this case, the p.d.f. of an observed process increment is a conditional density under the assumption that the process does not cross the degradation level in the current time interval. If the sample includes both process increments and failure times, then it is possible to estimate all parameters of the model.

These models and their parameter estimation are described in [KAH 04], [KAH 98] and in [KAH 10a]. A similar model and its application in medicine is described in [DOK 96]. Several generalizations of this simple model were given. It is possible to include measurement errors [WHI 95], or to transform the time scale [WHI 97]. It is also possible to consider bivariate models [WHI 98].

Here, we consider parameter estimation for the three different observation levels. In section 1.6.1, only degradation data are considered. We will find estimators for the parameters of a Wiener process with drift together with the two parameters describing the starting time and initial state of the process. Furthermore, confidence regions and improved confidence regions are

constructed. In the short section 1.6.2, we consider the classical sample of lifetimes. Finally, in section 1.6.3, we consider the case where, for each stochastic process, both degradation measures and a failure time are observable, and we estimate simultaneously the parameters μ, σ^2, t_0, x_0 and h. By contrast with the previous observations schemes, here, for each sample path of the degradation process, either a failure time is observed or degradation data at fixed observation times under the condition that the process has not yet exceeded the degradation level h. Hence, these conditional process increments and the failure time τ_h are dependent random variables. To compute the likelihood function, we have to find the conditional distribution of the process under the condition that the level h has not yet been exceeded, together with the joint distribution of conditional process increments and the lifetime τ_h.

1.6.1. *Statistical inference for degradation data*

In this section, we assume that the process is not stopped by a failure (which is equivalent to assuming that h tends to infinity). The observation of the underlying degradation process at fixed times gives an alternative approach to parameter estimation from lifetime data, which is particular useful for products with high reliability, since failure times have not to be observed. If we ignore the boundary level h, the parameters μ, σ^2, t_0 and x_0 may be estimated by use of data of the degradation process before a failure occurs. For this case, an example of a sample is given in Figure 1.7.

The sampling scheme is the following one: n i.i.d. copies of the degradation process $(X_t^{(1)})_{t\geq t_0}, \ldots, (X_t^{(n)})_{t\geq t_0}$ are observed. In addition, there are m_i observations of the i-th Wiener process at times $t_{i,1}, \ldots, t_{i,m_i}$. We now introduce several notations:

– $Y_{i,j} = X^{(i)}_{t_{i,j}} - X^{(i)}_{t_{i,j-1}}$: increments of the process;

– for $j \in \{2, \ldots, m_i\}$, $s_{i,j} = t_{i,j} - t_{i,j-1}$: delay between observation times;

– $X_{i,1} = X^{(i)}_{t_{i,1}}$: first observed degradation measure;

– $\bar{X}_{\cdot 1} = \frac{1}{n}\sum_{i=1}^{n} X^{(i)}_{t_{i,1}}$: average value of the first observed degradation measures;

– $\bar{X}_{\cdot m} = \frac{1}{n}\sum_{i=1}^{n} X^{(i)}_{t_{m_i}}$: average value of the last observed degradation measures;

– $\bar{t}_{\cdot 1} = \frac{1}{n}\sum_{i=1}^{n} t_{i,1}$: average value of the first observation times;

– $\bar{t}_{\cdot m} = \frac{1}{n}\sum_{i=1}^{n} t_{i,m_i}$: average value of the last observation times;

– $\bar{m} = \frac{1}{n}\sum_{i=1}^{n} m_i$: average value of the number of observations per process.

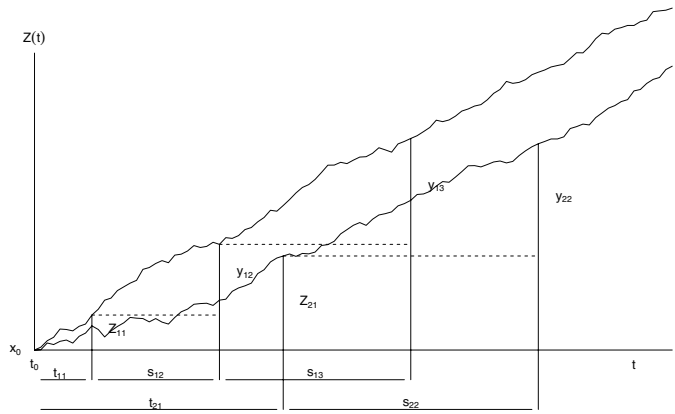

Figure 1.7. *Sample for observations of process increments for the case* $n = 2$, $m_1 = 3$, $m_2 = 2$

Lower case letter notations such as $y_{i,j}$ denote observed values of the corresponding random variables. Further, let $\mathbf{D}_{obs} = \bigcup_{i=1}^{n}\bigcup_{j=1}^{m_i}\{(t_{i,j}, X_{i,j})\}$ be the set of all observed degradation data. The increments are independent and Gaussian distributed:

$$X_{i,1} \sim \mathcal{N}(x_0 + \mu(t_{i,1} - t_0), \sigma^2(t_{i,1} - t_0)), \quad \forall i \in \{1, \ldots, n\}$$
$$Y_{i,j} \sim \mathcal{N}(\mu s_{i,j}, \sigma^2 s_{i,j}), \quad \forall j \in \{1, \ldots, m_i\}, \forall i \in \{1, \ldots, n\}$$

and the resulting likelihood function has the form:

$$L(x_0, t_0, \mu, \sigma^2 | \mathbf{D}_{obs}) = \prod_{i=1}^{n} \frac{1}{\sqrt{\sigma^2(t_{i,1} - t_0)}} \varphi\left(\frac{x_{i,1} - x_0 - \mu(t_{i,1} - t_0)}{\sqrt{\sigma^2(t_{i,1} - t_0)}}\right)$$
$$\times \prod_{i=1}^{n}\prod_{j=2}^{m_i} \frac{1}{\sqrt{\sigma^2 s_{i,j}}} \varphi\left(\frac{y_{i,j} - \mu s_{i,j}}{\sqrt{\sigma^2 s_{i,j}}}\right), \qquad [1.14]$$

where φ is p.d.f. of the standard normal distribution. Here, it has to be supposed that all the first observation times $t_{1,1}, \ldots, t_{n,1}$ are greater than t_0.

In this case, which is formally obtained on setting $h = \infty$, we do not make use of information on failure times nor of the fact that the degradation process does not exceed the level h between two successive observation times. The maximum likelihood equations are as follows:

$$\begin{cases} \widehat{\mu} = (\bar{x}_{\cdot m} - \widehat{x}_0)/(\bar{t}_{\cdot m} - \widehat{t}_0) \\ \widehat{\sigma}^2 = \dfrac{1}{\bar{m}n}\left(\displaystyle\sum_{i=1}^{n}\left(\frac{(x_{i,1}-\widehat{x}_0)^2}{t_{i,1}-\widehat{t}_0} + \sum_{j=2}^{m_i}\frac{y_{i,j}^2}{s_{i,j}}\right) - \widehat{\mu}n(\bar{x}_{\cdot m} - \widehat{x}_0)\right) \\ \widehat{x}_0 = \left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{x_{i,1}}{t_{i,1}-\widehat{t}_0} - \frac{\bar{x}_{\cdot m}}{\bar{t}_{\cdot m}-\widehat{t}_0}\right) \Big/ \left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{1}{t_{i,1}-\widehat{t}_0} - \frac{1}{\bar{t}_{\cdot m}-\widehat{t}_0}\right) \\ \widehat{\sigma}^2 \displaystyle\sum_{i=1}^{n}\frac{1}{t_{i,1}-\widehat{t}_0} = \sum_{i=1}^{n}\frac{(x_{i,1}-\widehat{x}_0)^2}{(t_{i,1}-\widehat{t}_0)^2} - \widehat{\mu}^2 n. \end{cases}$$

The last two equations, after replacing $\widehat{\mu}$ and $\widehat{\sigma}^2$ in the last equation, can be solved numerically with respect to $\widehat{x}_0$ and $\widehat{t}_0$. If these estimates are found, we get estimates $\widehat{\mu}$ and $\widehat{\sigma}^2$ for the two process parameters. An essential simplification arises if the first observation times are identical for all degradation process: $t_{i,1} = t_1,\ i \in \{1, \ldots, n\}$. Then, an explicit solution of the likelihood equations can be found:

$$\begin{cases} \widehat{\mu} = (\bar{x}_{\cdot m} - \bar{x}_{\cdot 1})/(\bar{t}_{\cdot m} - t_1) \\ \widehat{\sigma}^2 = \dfrac{1}{\bar{m}(n-1)}\left(\displaystyle\sum_{i=1}^{n}\sum_{j=2}^{m_i}\frac{y_{i,j}^2}{s_{i,j}} - n\frac{(\bar{x}_{\cdot m} - \bar{x}_{\cdot 1})^2}{\bar{t}_{\cdot m} - t_1}\right) \\ \widehat{x}_0 = \left(\bar{x}_{\cdot 1}(\bar{t}_{\cdot m} - \widehat{t}_0) - \bar{x}_{\cdot m}(t_1 - \widehat{t}_0)\right)/(\bar{t}_{\cdot m} - t_1) \\ \widehat{t}_0 = t_1 - \dfrac{1}{\widehat{\sigma}^2}\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_{i,1}^2 - \bar{x}_{\cdot 1}^2\right). \end{cases} \quad [1.15]$$

In the following, we will only consider the simpler case of equal first observation times. Then, the likelihood function has some "nice" properties. First, it can be proved, see [KAH 04], that the estimators given in equation [1.15] are asymptotically Gaussian distributed if:

– either, for estimating all parameters, the number of observed processes n tends to infinity;

– or, for estimating μ and σ^2, we observe only one process, but $\bar{m}$ and $(\bar{t}_{.m} - t_1)$ tend to infinity.

For the situation where the process is already running for a while, it may appear that only the parameters μ and σ^2 are of interest, while x_0 and t_0 (which describe the beginning of the degradation) are nuisance parameters. If the first observation times are identical for all degradation processes, then new nuisance parameters ν_1 and ν_2 can be found, which are independent of the parameters of interest μ and σ^2. This can be seen from the structure of the likelihood function given by equation [1.14]: x_0 and t_0 are present only in the first part. The reparameterization

$$\begin{cases} \nu_1 = x_0 + \mu(t_1 - t_0) \\ \nu_2 = \sigma^2(t_1 - t_0) \end{cases}$$

leads to the likelihood function

$$L(\mu, \sigma^2, \nu_1, \nu_2 | \mathbf{D}_{obs}) = \prod_{i=1}^{n} \frac{1}{\sqrt{\nu_2}} \varphi\left(\frac{x_{i,1} - \nu_1}{\sqrt{\nu_2}}\right) \prod_{i=1}^{n} \prod_{j=2}^{m_i} \times \frac{1}{\sqrt{\sigma^2 s_{i,j}}} \varphi\left(\frac{y_{i,j} - \mu s_{i,j}}{\sqrt{\sigma^2 s_{i,j}}}\right). \quad [1.16]$$

Note that the new parameters describe the expectation and the variance of the degradation process at the first observation time t_1. The likelihood function now consists of two independent parts: one part includes the nuisance parameters, ν_1 and ν_2, while the second part includes the parameters of interest, μ and σ^2. If we are interested in the process parameters, only the last part of the likelihood function must be taken into consideration. The information matrix of the parameter $\theta = (\mu, \sigma^2, x_0, t_0)^T$ can be found to be

$$I_\theta = n \begin{bmatrix} \frac{\bar{t}_{.m} - t_1}{\sigma^2} & 0 & 0 & 0 \\ 0 & \frac{\bar{m} - 1}{2\sigma^4} & 0 & 0 \\ 0 & 0 & \frac{1}{\nu_2} & 0 \\ 0 & 0 & & \frac{1}{2\nu_2^2} \end{bmatrix}. \quad [1.17]$$

All parameter estimators are asymptotically independent. The fact that the maximum likelihood estimator is asymptotically normally distributed can be used to construct confidence intervals for each parameter or simultaneous confidence regions for some of them. There are two possibilities for constructing asymptotic confidence regions. The first one is based on the Wald test statistic. An isoline of the k-dimensional Gaussian distribution, where k is the number of parameters, can be determined such that inside of this line there is $(1 - \alpha)$ of the probability mass of the distribution. The equation of this level line is given by the quadratic form [KAH 04]:

$$(\widehat{\theta} - \theta)^T I_\theta (\widehat{\theta} - \theta) < a^2, \qquad [1.18]$$

where $\widehat{\theta} = (\widehat{\mu}, \widehat{\sigma}^2, \widehat{x}_0, \widehat{t}_0)^T$ and where a^2 is the quantile of the χ^2-distribution with $k = 4$ degrees of freedom. The second possibility is based on the log-likelihood ratio statistic. It is well-known (see, e.g. [BAR 86] or [COR 87]) that under the same conditions of regularity as for asymptotic normality, the log-likelihood ratio

$$w = 2\left[\log(L(\widehat{\theta}|\mathbf{D}_{obs}) - \log(L(\theta_0|\mathbf{D}_{obs})\right]$$

converges in distribution to a χ^2-distribution with k degrees of freedom, where k is the number of parameters of interest, $\widehat{\theta}$ the maximum likelihood estimator of all parameters and θ_0 is the maximum likelihood estimator of the nuisance parameters, which depend on the parameters of interest. The advantage of the likelihood ratio statistic is that all nuisance parameters are excluded while they are present in the asymptotic variance of MLE. The rate of convergence of the log-likelihood ratio statistic to the χ^2-distribution is $O(n^{-1})$, where n is the sample size, and it may be improved to $O(n^{-3/2})$ by the Bartlett adjustment R if the fourth derivatives of the log-likelihood and their expectations exist. The log-likelihood ratio with Bartlett adjustment has the following form:

$$w' = w/(1 + R/p),$$

where p is the number of parameters of interest and R is the Bartlett adjustment that depends on the expectations of derivatives of the log-likelihood function and of the derivatives of these expectations. This adjustment is quite difficult to calculate in general. It becomes a little simpler if the estimators are asymptotically independent, because then

$\mathbb{E}(\frac{\partial \log L(\theta|\mathbf{D}_{obs})}{\partial\theta_i \partial\theta_j}) = 0$ for $i \neq j$. In [KAH 04], the Bartlett adjustment was found to be $R = 11/(6n(\bar{m}-1))$.

As a particular case, we here discuss a little the special case where $t_0 = 0 = x_0$. For any $i \in \{1, \ldots, n\}$ and any $j \in \{1, \ldots, m_i\}$, let $Y_{i,j} = X^{(i)}_{t_{i,j}} - X^{(i)}_{t_{i,j-1}}$ (with the convention that $t_{i,0} = 0$): $Y_{i,j}$ is normally distributed with mean $\mu s_{i,j}$ and variance $\sigma^2 s_{i,j}$, where $s_{i,j} = t_{i,j} - t_{i,j-1}$. From the independence of $Y_{i,j}$'s, it is straightforward to obtain the likelihood function and it leads to the following estimators:

$$\widehat{\mu} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{i,j}}{\sum_{i=1}^{n}\sum_{j=1}^{m_i} s_{i,j}} = \frac{\sum_{i=1}^{n} x^{(i)}_{t_{i,m_i}}}{\sum_{i=1}^{n} t_{i,m_i}} \quad \text{and}$$

$$\widehat{\sigma}^2 = \frac{1}{n\bar{m}} \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{(y_{i,j} - \widehat{\mu} s_{i,j})^2}{s_{i,j}}. \qquad [1.19]$$

Under the same condition as in the general case, these estimators are also asymptotically normal.

EXAMPLE 1.15.– To compare all possible confidence regions, we consider joint confidence estimators for the process parameters of interest μ and σ^2 for equal first observation times.

Based on the Wald statistic, the confidence region is an isoline of the two-dimensional Gaussian distribution:

$$\frac{(\widehat{\mu}-\mu)^2}{\sigma^2}(\bar{t}_{\cdot m} - t_1) + \frac{(\widehat{\sigma}^2 - \sigma^2)^2}{2\sigma^4}(\bar{m}-1) \leq \frac{a^2}{n}, \qquad [1.20]$$

where $a^2 = -2\log\alpha$ is the $(1-\alpha)$-quantile of the χ^2-distribution with two degrees of freedom. The asymptotic variances of the estimators contain the unknown parameter σ^2 and there are two possibilities for dealing with this issue. The first and most usual possibility is to replace the variance by its consistent estimate and equation [1.20] becomes:

$$\frac{(\widehat{\mu}-\mu)^2}{\widehat{\sigma}^2}(\bar{t}_{\cdot m} - t_1) + \frac{(\widehat{\sigma}^2 - \sigma^2)^2}{2\widehat{\sigma}^4}(\bar{m}-1) \leq \frac{a^2}{n}. \qquad [1.21]$$

The joint confidence region is an ellipse. A second possibility is to leave the unknown parameter within the asymptotic variance. Then, equation [1.20] is not an ellipse any more, and for small sample sizes the two regions are quite different.

The log-likelihood ratio is given by:

$$\frac{(\widehat{\mu}-\mu)^2}{\sigma^2}(\bar{t}_{\cdot m}-t_1)+\left(\frac{\widehat{\sigma}^2}{\sigma^2}-1-\log\left(\frac{\widehat{\sigma}^2}{\sigma^2}\right)\right)\geq-\frac{2\log(\alpha)}{n}. \quad [1.22]$$

For the Bartlett correction, the right-hand side of equation [1.22] gets the factor $1+11/(12n(\bar{m}-1))$, the confidence region becomes a little larger for small $n(\bar{m}-1)$. The correction factor tends to 1 if the sample size $n(\bar{m}-1)$ tends to infinity.

In Figure 1.8, the confidence regions for a relatively large sample size of 60 observations are plotted. We have simulated $n=10$ trajectories of the process with $t_0=x_0=0$, $\mu=1$, $\sigma^2=0.5$, $s_{i,j}=1$ and $\bar{m}=6$. The Wald statistic with estimated variance leads to an ellipse, it is the smallest area. The difference between likelihood ratio statistic with and without correction factor is very small for this sample size. The star presents the MLE and "+" is the true parameter.

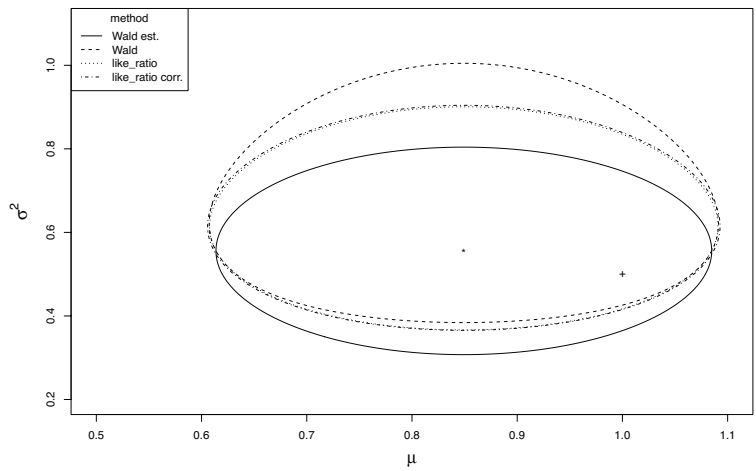

Figure 1.8. *Confidence regions for μ and σ^2 (60 observations)*

In Figure 1.9 are plotted the confidence regions for a relatively small sample size of 25 observations. We have simulated $n=5$ trajectories of the process with $t_0=x_0=0$, $\mu=1$, $\sigma^2=0.5$, $s_{i,j}=1$ and $\bar{m}=5$. Here, we have some difference between likelihood ratio statistic with and without correction factor. The Wald statistic leads to a very large confidence region.

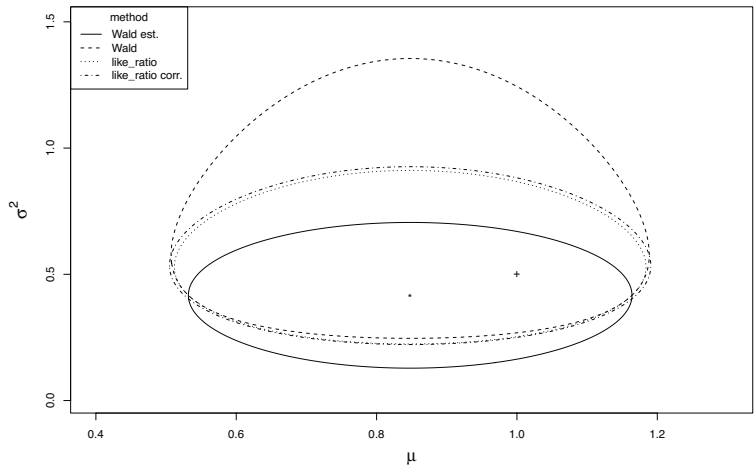

Figure 1.9. *Confidence regions for μ and σ^2 (25 observations)*

To compare the confidence level for finite sample sizes, a small simulation study is performed. We have simulated $n = 5$ trajectories of the process with $t_0 = x_0 = 0$, $\mu = 5$, $\sigma^2 = 0.25$, $s_{i,j} = 1$ and different $\bar{m}$. The number of observations $n\bar{m}$ goes from 15 ($\bar{m} = 3$) to 500 ($\bar{m} = 10$). For each sample size, we have simulated 1,000 samples, estimated the parameters and constructed the four confidence regions given above with confidence level $1-\alpha = 0.95$. We compute the empirical coverage probability. The results are given in Table 1.1.

$\bar{m}$	Equation [1.20]	Equation [1.21]	Equation [1.22]	Equation [1.22] corrected
3	95.40	83.93	93.95	94.82
5	95.19	87.65	94.12	94.68
7	95.27	89.50	94.44	94.74
10	95.51	91.92	95.16	95.43

Table 1.1. *Empirical confidence levels*

It can be seen that the best confidence regions (regarding confidence level), even for small sample sizes, are those based on the likelihood ratio statistic and Wald statistic without replacing the information by its estimate. The replacement, however, leads to a real confidence level smaller than the theoretical. The likelihood ratio statistic has for small sample sizes a smaller real confidence level than it should be, but it becomes better by introducing the Bartlett correction factor. Bearing in mind that the Wald statistic leads to

large confidence regions for small sample sizes, the likelihood ratio statistic is the best choice.

REMARK 1.16.– In the next sections, moment method estimators are considered as an alternative to MLE. For estimation based on observations of the Wiener process, the two methods coincide.

1.6.2. *Statistical inference for time to failure data*

When we know that a failure is the result of an underlying degradation process, the time to failure follows an inverse Gaussian distribution with p.d.f. given by equation [1.11]. Since the parameters of this time to failure are expressed with respect to the parameters of the underlying degradation process, it is possible to estimate the degradation parameters from a sample of failure times $\tau_1, \ldots, \tau_n$.

In this case, we get the classical likelihood function from a sample of inverse Gaussian distributed random variables:

$$L(x_0, t_0, \mu, \sigma^2 | \mathbf{D}_{obs}) = \prod_{i=1}^{n} \frac{h - x_0}{\sqrt{2\pi\sigma^2(\tau_i - t_0)^3}} \exp\left(-\frac{(h - x_0 - \mu(\tau_i - t_0))^2}{2\sigma^2(\tau_i - t_0)} \right) \mathbf{1}_{\{\tau_i > t_0\}}. \quad [1.23]$$

The estimators for the parameters μ and σ^2 can be calculated explicitly:

$$\widehat{\mu} = \frac{h - x_0}{\frac{1}{n}\sum_{i=1}^{n}(\tau_i - t_0)} \quad \text{and} \quad \widehat{\sigma}^2 = (h - x_0)^2 \frac{1}{n} \sum_{i=1}^{n} \frac{1}{(\tau_i - t_0)} - \widehat{\mu}(h - x_0).$$

Again, it can be shown that these estimators are asymptotically Gaussian distributed as n tends to infinity [KAH 04]. The problem of estimating t_0 is much more complicated, because it is a parameter that defines the support of the lifetime distribution. In [PAD 79], several methods were considered for estimating the parameters of the three-parameter inverse Gaussian distribution by maximum likelihood estimators, moment method estimators and a combination of these two methods.

Note that here we can only estimate the parameters μ, σ^2 and t_0 and that the degradation reserve $h - x_0$ must be given in advance.

1.6.3. *Statistical inference for both degradation and time to failure data*

Let $(X_t)_{t\geq t_0}$ be a degradation process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $(\mathbb{R}, \mathcal{B})$ as defined by equation [1.4], where $\mathcal{B}$ is the Borel σ-algebra on $\mathbb{R}$. Furthermore, let $t_1, \ldots, t_m$ with $m \in \mathbb{N}^*$ be fixed observation times with $t_0 < t_1 < \ldots < t_m < \infty$. To alleviate the notation, we consider only one single degradation process $(X_t)_{t\geq t_0}$ and thus we drop the subscript i. We assume that a failure is observable at any time $t > t_0$ and that we stop observing the degradation process as soon as a failure has occurred. Hence, in each time interval $(t_{j-1}, t_j]$, with $j \in \{1, \ldots, m\}$, we observe either a failure at time $\tau_h \in (t_{j-1}, t_j]$ or we observe the degradation measure $x_j = X_{t_j}$ at time t_j under the condition that the process has not yet exceeded the level h until the time t_j. The observable stopped process is given, for any $t \geq t_0$, by:

$$\widetilde{X}_t = \begin{cases} X_t & \text{if } t_0 \leq t \leq \tau_h, \\ \infty & \text{if } t > \tau_h. \end{cases}$$

Thus, $\widetilde{X}_t$ is a functional of $\{X(s); t_0 \leq s \leq t\}$:

$$\forall t \geq t_0, \quad \widetilde{X}_t = G_t(\tau_h, X_t), \tag{1.24}$$

where, for any $t, \tau \geq t_0$ and any $x \in \mathbb{R}$,

$$G_t(\tau, x) = \begin{cases} x & \text{if } \tau \geq t \\ \infty & \text{if } \tau < t \end{cases}$$

and $\widetilde{X}_t$ is measurable with respect to the σ-algebra $\mathcal{F}_{\mathrm{t}} = \sigma\{X_s;\ t_0 \leq s \leq t\}$.

A sample of censored observations has the structure:

$$X = F(\tau_h, X_{t_1}, \ldots, X_{t_m}) = (\min(\tau_h, t_m), \widetilde{X}_{t_1}, \ldots, \widetilde{X}_{t_m}) \tag{1.25}$$

with

$$F(\tau, x_1, \ldots, x_m) = (F_0(\tau), G_{t_1}(\tau, x_1), \ldots, G_{t_m}(\tau, x_m))$$

and

$$F_0(\tau) = \min(\tau, t_m).$$

The statistical model is now given by:

$$(\mathbb{R} \times \overline{\mathbb{R}}^m, \mathcal{B} \otimes \overline{\mathcal{B}}^m, (\mathbb{P}_\theta^F)_{\theta=(\mu,\sigma^2,x_0,t_0,h)\in\Theta\subset\mathbb{R}^5}),$$

where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$, $\overline{\mathcal{B}}$ is the σ-algebra of Borel sets in $\overline{\mathbb{R}}$ and $\mathbb{P}_\theta^F$ is the distribution of F when the model parameter is θ.

To compute the likelihood function of X, we have to find the Radon–Nikodym derivative of $\mathbb{P}_\theta^F$ with respect to a dominating measure ν:

$$L(x_0, t_0, \mu, \sigma^2 | \mathbf{D}_{obs}) = \frac{\mathrm{d}\mathbb{P}_\theta^F}{\mathrm{d}\nu}(\tau, \mathbf{x}),\ \tau \in \mathbb{R}, \mathbf{x} = (x_1, \ldots, x_m) \in \overline{\mathbb{R}}^m. \quad [1.26]$$

A crucial step in the derivation of $L(x_0, t_0, \mu, \sigma^2 | \mathbf{D}_{obs})$ is the calculation of the conditional density $z \mapsto f(s, x, t, z, h)$ of the process X_t at time t starting at point $x = X_s$ under the condition that the level h is not exceeded in $[s, t]$. For that purpose, for $s \geq t_0$ and $x \in \mathbb{R}$, let $\mathbb{P}_{s,x}$ stand for the conditional probability $\mathbb{P}_{s,x} = \mathbb{P}(\cdot | X_s = x)$ and let us introduce the transition kernel $(\mathbb{P}(s, x, t, \mathrm{d}y))_{t \geq t_0}$ of the Markov process $(X_t)_{t \geq t_0}$:

$$\begin{aligned}\mathbb{P}(s, x, t, B) &= \mathbb{P}(X_t \in B | X_s = x) \\ &= \frac{1}{\sqrt{2\pi\sigma^2(t-s)}} \int_B \exp\left(-\frac{(y - x - \mu(t-s))^2}{2\sigma^2(t-s)}\right) \mathrm{d}y\end{aligned}$$

for all $t > s \geq t_0$, all $x \in \mathbb{R}$ and all $B \in \mathcal{B}$.

$\mathbb{P}_{s,x}$ is a probability measure on $\mathcal{F}_{\geq s} = \sigma\{X_u;\ u \geq s\}$. Let τ_h^s denote the first passage time of X_t to the upper boundary h on $[s, \infty)$:

$$\tau_h^s = \inf\{t \geq s; X_t \geq h\}.$$

In the case where $s = t_0$, we have $\tau_h^{t_0} = \tau_h$ according to equation [1.6]. For every fixed $s \geq t_0$ and $x \in \mathbb{R}$, τ_h^s is a random variable on the probability space $(\Omega, \mathcal{F}_{\geq s}, \mathbb{P}_{s,x})$. Using a similar argument as for the distribution of τ_h

in [1.11], for $x < h$, the induced probability measure $\mathbb{P}_{s,x}^{\tau_h^s}$, that is the conditional distribution of τ_h^s under the condition $\{X_s = x\}$, has the following p.d.f.:

$$g(s,x,t,h) = \frac{h-x}{\sqrt{2\pi\sigma^2(t-s)^3}} \exp\left(-\frac{(h-x-\mu(t-s))^2}{2\sigma^2(t-s)}\right) \mathbf{1}_{\{t>s\}}. \quad [1.27]$$

If $x \geq h$, then $\mathbb{P}_{s,x}^{\tau_h^s}$ is the Dirac measure δ_s concentrated at s. For $x < h$, $z \in \mathbb{R}$ and $s < t$, writing

$$\mathbb{P}_{s,x}(X_t \leq z, \tau_h^s > t) = \mathbb{P}_{s,x}(X_t \leq z) - \mathbb{P}_{s,x}(X_t \leq z, \tau_h^s \leq t)$$

and using mostly similar arguments as for [1.7] (total probability law plus the Markov property), we get:

$$\begin{aligned}
&\mathbb{P}_{s,x}(X_t \leq z, \tau_h^s > t) \\
&\quad = \mathbb{P}_{s,x}(X_t \leq z) - \int_s^t \mathbb{P}_{s,x}(X_t \leq z \,|\, \tau_h^s = u)\, \mathbb{P}_{s,x}^{\tau_h^s}(\mathrm{d}u) \\
&\quad = \mathbb{P}(s,x,t,(-\infty,z]) - \int_s^t \mathbb{P}(u,h,t,(-\infty,z])\, g(s,x,u,h)\, \mathrm{d}u
\end{aligned}$$

and hence, applying Fubini's theorem and passing to densities,

$$f(s,x,t,z,h) = p(s,x,t,z) - \int_s^t p(u,h,t,z)\, g(s,x,u,h)\, \mathrm{d}u, \quad [1.28]$$

where $p(s,x,t,z)$ is the transition density of $(X_t, \mathbb{P}_{s,x})$ (namely $\mathbb{P}(s,x,t,\mathrm{d}z) = p(s,x,t,z)\, \mathrm{d}z$). From equation [1.28], the following lemma provides an explicit representation of the conditional density of a process increment under the condition that the process does not cross the level h:

$$f(s,x,t,z,h) = \frac{\partial}{\partial z}\mathbb{P}(X_t \leq z, \tau_h \notin [s,t) | X_s = x).$$

LEMMA 1.17.– Let $t_0 \leq s < t, x < h$ and $z \in \mathbb{R}$. The p.d.f. in equation [1.28] is given by

$$f(s,x,t,z,h) = \frac{1}{\sigma\sqrt{t-s}}\varphi\left(\frac{z-x-\mu(t-s)}{\sigma\sqrt{t-s}}\right) \left[1-\exp\left(-\frac{2(h-x)(h-z)}{\sigma^2(t-s)}\right)\right]\mathbf{1}_{\{z\leq h\}}, \quad [1.29]$$

where φ is the p.d.f. of the standard normal distribution.

PROOF.– Recalling equation [1.28], we have to compute the following integral:

$$\begin{aligned} I &= \int_s^t p(u,h,t,z)\, g(s,x,u,h)\, \mathrm{d}u \\ &= \int_s^t \frac{h-x}{2\pi\sigma^2\sqrt{(t-u)(u-s)^3}} \\ &\quad \times \exp\left[-\frac{1}{2\sigma^2}\left(\frac{(z-h-\mu(t-u))^2}{t-u} + \frac{(h-x-\mu(u-s))^2}{u-s}\right)\right] \mathrm{d}u \end{aligned}$$

for $t > s \geq t_0$ and $x < h$. Applying the substitution, $y = \sqrt{(t-u)/(u-s)}$ for $s < u < t$, i.e. $u = (t+sy^2)/(1+y^2)$ with $+\infty > y > 0$, we obtain after some algebra that

$$\begin{aligned} I &= \frac{h-x}{\pi\sigma^2(t-s)} \\ &\quad \times \exp\left[-\frac{(z-h)^2+(h-x)^2-2\mu(t-s)(z-x)+\mu^2(t-s)^2}{2\sigma^2(t-s)}\right] \\ &\quad \times \int_0^\infty \exp\left[-\left(\frac{\beta^2}{y^2}+\alpha^2y^2\right)\right] \mathrm{d}y \end{aligned}$$

with $\beta^2 = (z-h)^2/(2\sigma^2(t-s))$ and $\alpha^2 = (h-x)^2/(2\sigma^2(t-s))$. Denoting the last integral by I_1, it can be easily checked that for $\alpha \neq 0$:

$$\begin{aligned} I_1 &= \frac{\sqrt{\pi}}{4\alpha}\left[\mathrm{e}^{2\alpha\beta}\left(\mathrm{erf}\left(\alpha z+\frac{\beta}{z}\right)-1\right)+\mathrm{e}^{-2\alpha\beta}\left(\mathrm{erf}\left(\alpha z-\frac{\beta}{z}\right)+1\right)\right]\Bigg|_{z=0}^{z=\infty} \\ &= \frac{\sqrt{\pi}}{2|\alpha|}\,\mathrm{e}^{-2|\alpha\beta|}, \end{aligned}$$

where $\mathrm{erf}(x) = 2\pi^{-1/2}\int_0^x \mathrm{e}^{-t^2}\,\mathrm{d}t$ is the error function. Hence, as $x < h$ and $|\alpha| = |h-x|(2\sigma^2(t-s))^{-1/2}$, we get

$$I = \frac{1}{\sqrt{2\pi\sigma^2(t-s)}} \exp\left[-\frac{Q - 2\mu(t-s)(z-x) + \mu^2(t-s)^2}{2\sigma^2(t-s)}\right]$$

with

$$\begin{aligned} Q &= (z-h)^2 + 2|z-h|(h-x) + (h-x)^2 \\ &= \begin{cases} (z-2h+x)^2 = (z-x)^2 + 4(h-x)(h-z) & \text{if } z \le h \\ (z-x)^2 & \text{if } z > h \end{cases} \end{aligned}$$

and, finally,

$$I = \frac{1}{\sigma\sqrt{t-s}}\varphi\left(\frac{z-x-\mu(t-s)}{\sigma\sqrt{t-s}}\right)\begin{cases} \exp\left(-\dfrac{2(h-x)(h-z)}{\sigma^2(t-s)}\right) & \text{if } z \le h \\ 1 & \text{if } z > h \end{cases},$$

where φ is the density of the standard normal distribution. This completes the proof. □

The next lemma [KAH 10a] gives, in terms of f the density of the finite dimensional distribution of X_t starting at $x = X_s$ under the condition that h is not exceeded. Let λ^k be the k-dimensional Lebesgue measure on $\overline{\mathcal{B}}^{\mathrm{k}}$ ($k \in \mathbb{N}$). We use the convention $\prod_{j=j_1}^{j_2} \alpha_j = 1$ if $j_2 < j_1$.

LEMMA 1.18.– Let $t_0 \le s < t_1$, $\mathbf{t}_{1,k} = (t_1, \ldots, t_k)$, $\mathbf{x}_{1,k} = (x_1, \ldots, x_k) \in \mathbb{R}^k$ and $\mathbf{X}_{\mathbf{t}_{1,k}} = (X_{t_1}, \ldots, X_{t_k})$, for $1 \le k \le m$. Then, the measure $\mathbb{P}_{s,x}\left(\mathbf{X}_{\mathbf{t}_{1,k}} \in \cdot, \tau_h^s \ge t_k\right)$ has the following density with respect to Lebesgue measure λ^k:

$$f_k(s, x, \mathbf{t}_{1,k}, \mathbf{x}_{1,k}, h) = f(s, x, t_1, x_1, h) \prod_{j=2}^{k} f(t_{j-1}, x_{j-1}, t_j, x_j, h). \quad [1.30]$$

PROOF.– For $k = 1$, clearly $f_1(s, x, \mathbf{t}_{1,1}, \mathbf{x}_{1,1}, h) = f(s, x, t_1, x_1, h)$ by the definition of f. For simplicity, we will prove equation [1.30] only for $k = 2$. The general case can be shown analogously by induction. Let $B_1, B_2 \in \mathcal{B}$. Then, applying the law of total probability and using the Markov property, we

have

$$\begin{aligned}
&\mathbb{P}_{s,x}(\mathbf{X}_{\mathbf{t}_{1,2}} \in B_1 \times B_2,\, \tau_h^s \geq t_2) \\
&\quad = \mathbb{P}_{s,x}(X_{t_2} \in B_2,\, \tau_h^{t_1} \geq t_2,\, X_{t_1} \in B_1,\, \tau_h^s \geq t_1) \\
&\quad = \int_{B_1} \mathbb{P}_{s,x}(X_{t_2} \in B_2,\, \tau_h^{t_1} \geq t_2 \,|\, X_{t_1} = x_1)\, f(s,x,t_1,x_1,h)\, \mathrm{d}x_1 \\
&\quad = \int_{B_1} \mathbb{P}_{t_1,x_1}(X_{t_2} \in B_2,\, \tau_h^{t_1} \geq t_2)\, f(s,x,t_1,x_1,h)\, \mathrm{d}x_1 \\
&\quad = \int_{B_2 \times B_1} f(t_1,x_1,t_2,x_2,h)\, f(s,x,t_1,x_1,h)\, \mathrm{d}x_2\, \mathrm{d}x_1.
\end{aligned}$$

□

Based on these preliminary lemmas, we can derive the likelihood function $L(x_0, t_0, \mu, \sigma^2|\mathbf{D}_{obs})$ [KAH 10a]. We shall show that $\mathbb{P}_\theta^F$ is, for all $\theta \in \Theta$, absolutely continuous with respect to the measure

$$\nu = (\lambda + \delta_{t_m}) \otimes (\lambda + \delta_\infty)^m, \tag{1.31}$$

where λ is the Lebesgue measure on $\overline{\mathcal{B}}$ and, for every $x \in \overline{\mathbb{R}}$, δ_x is the Dirac measure on $\overline{\mathcal{B}}$ concentrated at x.

THEOREM 1.19.– Let ν be defined as in equation [1.31] on $\mathcal{B} \otimes \overline{\mathcal{B}}^{\mathrm{m}}$. Then, for all $\theta \in \Theta$, $\mathbb{P}_\theta^F$ is absolutely continuous with respect to ν and has the following Radon–Nikodym derivative:

$$\begin{aligned}
&L(x_0, t_0, \mu, \sigma^2|\mathbf{D}_{obs}) = \frac{\mathrm{d}\mathbb{P}_\theta^F}{\mathrm{d}\nu}(\tau, \mathbf{x}) \\
&= \prod_{k=1}^{m} \left(g(t_{k-1}, x_{k-1}, \tau, h) \prod_{j=1}^{k-1} f(t_{j-1}, x_{j-1}, t_j, x_j, h) \right. \\
&\qquad \left. \times \prod_{j=k}^{m} \mathbf{1}_{\{x_j = \infty\}} \right)^{\mathbf{1}_{\{t_{k-1} \leq \tau < t_k\}}} \\
&\qquad \times \left(\prod_{j=1}^{m} f(t_{j-1}, x_{j-1}, t_j, x_j, h) \right)^{\mathbf{1}_{\{\tau = t_m\}}} \mathbf{1}_{\{t_0 \leq \tau \leq t_m\}},
\end{aligned} \tag{1.32}$$

where $\tau \in \mathbb{R}$, and $\mathbf{x} = (x_1, \ldots, x_m) \in \overline{\mathbb{R}}^m$.

PROOF.– Let $1 \leq i \leq j \leq n$, $(y_i, \ldots, y_j) \in \overline{\mathbb{R}}^{j-i+1}$ and $t \in \mathbb{R}$. For the sake of brevity, we define $\mathbf{t}_{i,j} = (t_i, \ldots, t_j)$, $\mathbf{y}_{i,j} = (y_i, \ldots, y_j)$, $(-\infty, \mathbf{y}_{i,j}] = (-\infty, y_i] \times \cdots \times (-\infty, y_j]$ and $E = (-\infty, t] \times (-\infty, \mathbf{y}_{1,m}]$. Furthermore, we use the abbreviation $\{\mathbf{X}_{\mathbf{t}_{i,j}} \leq \mathbf{y}_{i,j}\} = \{X_{t_i} \leq y_i, \ldots, X_{t_j} \leq y_j\}$ and an analogous one for $\widetilde{\mathbf{X}}_t$. From equation [1.25], we have, for all $\theta \in \Theta$:

$$\begin{aligned}\mathbb{P}_\theta^F(E) &= \mathbb{P}_{t_0,x_0}(\min(\tau_h, t_m) \leq t, \widetilde{\mathbf{X}}_{\mathbf{t}_{1,m}} \leq \mathbf{y}_{1,m}) \\ &= \sum_{k=1}^{m} \mathbf{1}_{\{t \geq t_{k-1}\}} \mathbb{P}_{t_0,x_0}(t_{k-1} \leq \tau_h \leq \min(t, t_k), \widetilde{\mathbf{X}}_{\mathbf{t}_{1,m}} \leq \mathbf{y}_{1,m}) \\ &\qquad + \mathbf{1}_{\{t \geq t_m\}} \mathbb{P}_{t_0,x_0}(\tau_h \geq t_m,\ \widetilde{\mathbf{X}}_{\mathbf{t}_{1,m}} \leq \mathbf{y}_{1,m}). \end{aligned} \quad [1.33]$$

To calculate the first m terms of equation [1.33], take $t_{k-1} \leq t < t_k$ with $1 \leq k \leq m$ and define $\bigcap_{k=k_1}^{k_2} A_k = \Omega$ if $k_2 < k_1$. Then, for $s \geq t_{k-1}$, by the definition of $\widetilde{\mathbf{X}}_s$:

$$\begin{aligned}&\mathbb{P}_{t_0,x_0}(t_{k-1} \leq \tau_h \leq s, \widetilde{\mathbf{X}}_{\mathbf{t}_{1,m}} \leq \mathbf{y}_{1,m}) \\ &= \mathbb{P}_{t_0,x_0}(\{t_{k-1} \leq \tau_h \leq s,\ \mathbf{X}_{\mathbf{t}_{1,k-1}} \leq \mathbf{y}_{1,k-1}\} \cap \{\widetilde{\mathbf{X}}_{\mathbf{t}_{k,m}} \leq \mathbf{y}_{k,m}\}) \\ &= \mathbb{P}_{t_0,x_0}(\tau_h^{t_{k-1}} \leq s,\ \tau_h^{t_0} \geq t_{k-1}, \mathbf{X}_{\mathbf{t}_{1,k-1}} \leq \mathbf{y}_{1,k-1}) \prod_{j=k}^{m} \mathbf{1}_{\{y_j = \infty\}}.\end{aligned}$$

Applying the law of total probability, the Markov property and lemma 1.18, it yields

$$\begin{aligned}&\mathbb{P}_{t_0,x_0}(\tau_h^{t_{k-1}} \leq s,\ \tau_h^{t_0} \geq t_{k-1}, \mathbf{X}_{\mathbf{t}_{1,k-1}} \leq \mathbf{y}_{1,k-1}) \\ &= \int_{(-\infty, \mathbf{y}_{1,k-1}]} \mathbb{P}_{t_0,x_0}(\tau_h^{t_{k-1}} \leq s \,|\, X_{t_{k-1}} = x_{k-1}) \\ &\qquad \times f_{k-1}(t_0, x_0, \mathbf{t}_{1,k-1}, \mathbf{x}_{1,k-1}, h) \lambda^{k-1}(\mathrm{d}\mathbf{x}_{1,k-1}).\end{aligned}$$

Moreover, recalling the distribution of $\mathbb{P}_{s,x}^{\tau_h^s}$, we have:

$$\begin{aligned}\mathbb{P}_{t_0,x_0}(\tau_h^{t_{k-1}} \leq s \,|\, X_{s_{k-1}} = x_{k-1}) &= \mathbb{P}_{t_{k-1},x_{k-1}}(\tau_h^{t_{k-1}} \leq s) \\ &= \int_{t_{k-1}}^{s} g(t_{k-1}, x_{k-1}, \tau, h)\, \lambda(\mathrm{d}\tau)\end{aligned}$$

if $x_{k-1} < h$. Finally, the product of indicator functions in equation [1.19] can be written as:

$$\prod_{j=k}^{m} \mathbf{1}_{\{y_j=\infty\}} = \int_{(-\infty,\, \mathbf{y}_{k,m}]} \prod_{j=k}^{m} \mathbf{1}_{\{x_j=\infty\}} \delta_\infty^{m-k+1}(\mathrm{d}\mathbf{x}_{k,m}).$$

Combining the above equations, we obtain:

$$\begin{aligned}
&\mathbb{P}_{t_0,x_0}(t_{k-1} \leq \tau_h \leq s, \widetilde{\mathbf{X}}_{\mathbf{t}_{1,m}} \leq \mathbf{y}_{1,m}) \\
&= \int_{(-\infty,\, \mathbf{y}_{1,m}]} \int_{t_{k-1}}^{s} g(t_{k-1}, x_{k-1}, \tau, h)\, \lambda(\mathrm{d}\tau)\; f_{k-1}(t_0, x_0, \mathbf{t}_{1,k-1}, \mathbf{x}_{1,k-1}, h) \\
&\qquad \times \prod_{j=k}^{m} \mathbf{1}_{\{x_j=\infty\}} (\lambda^{k-1} \otimes \delta_\infty^{m-k+1})(\mathrm{d}\mathbf{x}_{1,m}). \quad [1.34]
\end{aligned}$$

Now, from equations [1.29] and [1.30], it is obvious that the measure $\lambda^{k-1} \otimes \delta_\infty^{m-k+1}$ given in equation [1.34] can be replaced by:

$$(\lambda + \delta_\infty)^m = \sum_{(\mu_1,\ldots,\mu_m) \in \{\lambda,\, \delta_\infty\}^m} \bigotimes_{i=1}^{m} \mu_i$$

without altering the value of the integral, since integrating with respect to the summands of $(\lambda+\delta_\infty)^m$ gives zero except for the case $\mu_i = \lambda^{k-1} \otimes \delta_\infty^{m-k+1}$. Hence, going back to equation [1.33], using equation [1.19] and lemma 1.18, we obtain:

$$\begin{aligned}
\mathbb{P}_\theta^F(E) &= \int_{(-\infty,\, \mathbf{y}_{1,m}]} \left(\int_{t_0}^{t} \sum_{k=1}^{m} \mathbf{1}_{\{t_{k-1} \leq \tau < t_k\}} g(t_{k-1}, x_{k-1}, \tau, h)\, \lambda(\mathrm{d}\tau) \right. \\
&\quad \times f_{k-1}(t_0, x_0, \mathbf{t}_{1,k-1}, \mathbf{x}_{1,k-1}, h) \prod_{j=k}^{m} \mathbf{1}_{\{x_j=\infty\}} \\
&\quad \left. + \int_{t_0}^{t} \mathbf{1}_{\{\tau=t_m\}}\, \delta_{t_m}(\mathrm{d}\tau)\; f_m(t_0, x_0, \mathbf{t}_{1,m}, \mathbf{x}_{1,m}, h) \right) (\lambda + \delta_\infty)^m(\mathrm{d}\mathbf{x}_{1,m}) \\
&= \int_{(-\infty,t] \times (-\infty,\, \mathbf{y}_{1,m}]} \left(\sum_{k=1}^{m} \mathbf{1}_{\{t_{k-1} \leq \tau < t_k\}} g(t_{k-1}, x_{k-1}, \tau, h) \right.
\end{aligned}$$

$$\times \prod_{j=1}^{k-1} f(t_{j-1}, x_{j-1}, t_j, x_j, h) \prod_{j=k}^{m} \mathbf{1}_{\{x_j=\infty\}}$$

$$+ \mathbf{1}_{\{\tau=t_m\}} \prod_{j=1}^{m} f(t_{j-1}, x_{j-1}, t_j, x_j, h) \Bigg) \nu(\mathrm{d}(\tau, \mathbf{x}_{1,m})),$$

where $\nu = (\lambda + \delta_{t_m}) \otimes (\lambda + \delta_\infty)^m$. Thus, the integrand of the last equation is a version of the Radon–Nikodym derivative of $\mathbb{P}_\theta^F$ with respect to ν, since $\mathcal{B} \otimes \overline{\mathcal{B}}^{\mathrm{m}}$ is generated by the sets E. To complete the proof note that this integrand and the right-hand side of equation [1.32] are equivalent. □

In general, we observe not only one single degradation process, but n i.i.d. degradation processes $X^{(1)}, \ldots, X^{(n)}$ corresponding to n independent and identical items. Let $t_{i1}, \ldots, t_{im_i}$ with $t_0 < t_{i1} < \ldots < t_{im_i} < \infty$ be the m_i fixed observation times for $X^{(i)}$, $M = \sum_{i=1}^n m_i$ and $\tau_h^{(i)}$, the lifetime associated with $X^{(i)}$. In this case, the censored observations have the form:

$$\begin{aligned}&(\tau_i, x_{i1}, \ldots, x_{im_i})\\ =&(\min(\tau_h^{(i)}, t_{im_i}), G_{t_{i1}}(\tau_h^{(i)}, X^{(i)}(t_{i1})), \ldots, G_{t_{im_i}}(\tau_h^{(i)}, X^{(i)}(t_{im_i}))).\end{aligned}$$

Then, the statistical model is given by:

$$(\mathbb{R}^n \times \overline{\mathbb{R}}^M, \mathcal{B}^n \otimes \overline{\mathcal{B}}^M, (\otimes_{i=1}^n \mathbb{P}_\theta^F)_{\theta=(\mu,\sigma^2,x_0,t_0,h)\in\Theta\subset\mathbb{R}^5})$$

with the following likelihood function:

$$L(x_0, t_0, \mu, \sigma^2 | \mathbf{D}_{obs}) = \prod_{i=1}^{n} L^{(i)}(x_0, t_0, \mu, \sigma^2 | \mathbf{D}_{obs}) \ ,$$

where $\tau = (\tau_1, \ldots, \tau_n) \in \mathbb{R}^n$, $\mathbf{x}^{(i)} = (x_{i1}, \ldots, x_{im_i}) \in \overline{\mathbb{R}}^{m_i}$, $i \in \{1, \ldots, n\}$ and $L^{(i)}$ is the likelihood function of the i-th observed process.

EXAMPLE 1.20.– We consider degradation processes with $x_0 = 0$ and $t_0 = 0$. Substituting equations [1.27] and [1.29] for the functions g and f in equation [1.32], we get the likelihood function of the i-th observed process

for $i \in \{1, \ldots, n\}$,

$$L(\mu, \sigma^2, h|\mathbf{D}_{obs}) = \prod_{j=1}^{k-1} \left(\frac{1}{\sqrt{\sigma^2(t_{i,j} - t_{i,j-1})}} \varphi \left(\frac{(x_{i,j} - x_{i,j-1}) - \mu(t_{i,j} - t_{i,j-1})}{\sqrt{\sigma^2(t_{i,j} - t_{i,j-1})}} \right) \times \left[1 - \exp\left(-\frac{2(h - x_{i,j-1})(h - x_{i,j})}{\sigma^2(t_{i,j} - t_{i,j-1})} \right) \right] \right) \prod_{j=k}^{m_i} \mathbf{1}_{\{x_j = \infty\}} \times \frac{h - x_{i,k-1}}{\sqrt{\sigma^2(\tau_i - t_{i,k-1})^3}} \varphi \left(\frac{h - x_{i,k-1} - \mu(\tau_i - t_{i,k-1})}{\sqrt{\sigma^2(\tau_i - t_{i,k-1})}} \right), \quad [1.35]$$

if $t_{k-1} \leq \tau_i < t_k$ and $1 \leq k \leq m_i$ and

$$L(\mu, \sigma^2, h|\mathbf{D}_{obs}) = \prod_{j=1}^{m_i} \left(\frac{1}{\sqrt{\sigma^2(t_{i,j} - t_{i,j-1})}} \varphi \left(\frac{(x_{i,j} - x_{i,j-1}) - \mu(t_{i,j} - t_{i,j-1})}{\sqrt{\sigma^2(t_{i,j} - t_{i,j-1})}} \right) \times \left[1 - \exp\left(-\frac{2(h - x_{i,j-1})(h - x_{i,j})}{\sigma^2(t_{i,j} - t_{i,j-1})} \right) \right] \right), \quad [1.36]$$

if $\tau_i \geq t_{m_i}$. To compare our estimates based on the likelihood function given by equation [1.35] and by equation [1.36] with previous results, we consider the case where only degradation process increments are observed. As shown in section 1.6.1, in this case, the likelihood function of the i-th observed process is given by:

$$L_D(\mu, \sigma^2|\mathbf{D}_{obs}) = \prod_{j=1}^{m_i} \frac{1}{\sqrt{\sigma^2(t_{i,j} - t_{i,j-1})}} \varphi \left(\frac{(x_{i,j} - x_{i,j-1}) - \mu(t_{i,j} - t_{i,j-1})}{\sqrt{\sigma^2(t_{i,j} - t_{i,j-1})}} \right). \quad [1.37]$$

We notice that the likelihood function given by equation [1.37] can be obtained from equation [1.36] by making h tend to infinity, which means that the boundary level h does not exist and, consequently, exceeding the boundary is impossible.

As an example, we have simulated a sample of $n = 5$ Wiener processes with $m_i = 10$ observation times $t_{i,j}$ for each process, which have equal

distances $t_{i,j} - t_{i,j-1} = 1$ and with true parameter values $\mu = 1.0$, $\sigma^2 = 0.25$, $t_0 = 0$, $x_0 = 0$ and $h = 5.0$. The simulated data are shown in Table 1.2.

i	1	2	3	4	5
$x_{i,j} - x_{i,j-1}$	1.290	0.397	0.744	0.984	1.223
	1.096	1.619	1.125	0.904	0.769
	0.715	0.427	1.047	1.384	1.775
	0.339	1.574	0.530	1.000	0.567
	0.028		0.123		0.426
	1.231		1.346		
τ_i	6.559	4.967	6.018	4.461	5.138

Table 1.2. *A simulated sample*

Supposing that the parameter values $t_0 = 0$ and $x_0 = 0$ are known, we first estimate the parameters μ and σ^2 maximizing the likelihood function given by equation [1.4], which means that we ignore that the degradation process does not exceed the boundary level between observation times. Using the maximum likelihood estimators described in section 1.6.1, we get:

$$\widehat{\mu}_D = 0.9198, \quad \text{and} \quad \widehat{\sigma}^2_D = 0.1983.$$

In Figure 1.10, 95 and 99% confidence regions for μ and σ^2 based on the asymptotic χ^2 distribution (with two degrees of freedom) of the likelihood ratio are shown for this model.

Now, supposing again that $t_0 = 0$ and $x_0 = 0$ are known, we compute the maximum likelihood estimates for the parameters μ, σ^2 and h based on the likelihood function given by equation [1.35] of the exact model. The maximization of the likelihood function yields the point estimates

$$\widehat{\mu} = 0.9222, \quad \widehat{\sigma^2} = 0.1915 \quad \text{and} \quad \widehat{h} = 5.0047.$$

The numerical optimization procedure turned out to be difficult to converge, so that a good initial value of the parameters has to be known. Such initial values may be the estimates of μ and σ^2 from the likelihood function given by equation [1.4] and an initial value of h can be found from the likelihood function based on the inverse Gaussian density (see equation [1.11]). Figures 1.11– 1.13 show 0.95- and 0.99-confidence regions based on the asymptotic χ^2_2 distribution of the likelihood ratio for the parameters (μ, σ^2), (μ, h) and (σ^2, h), respectively.

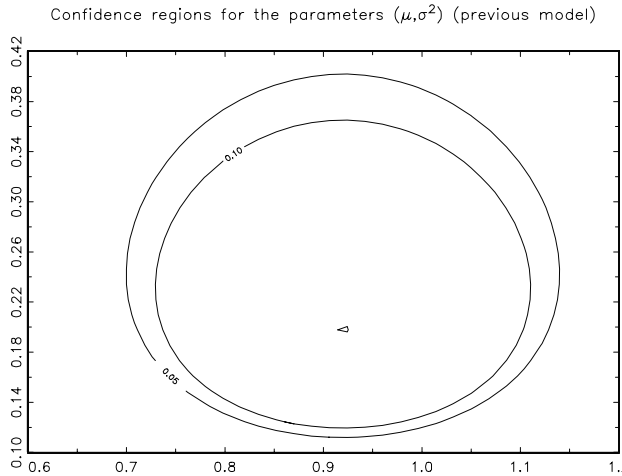

Figure 1.10. *Confidence estimates for* (μ, σ^2)

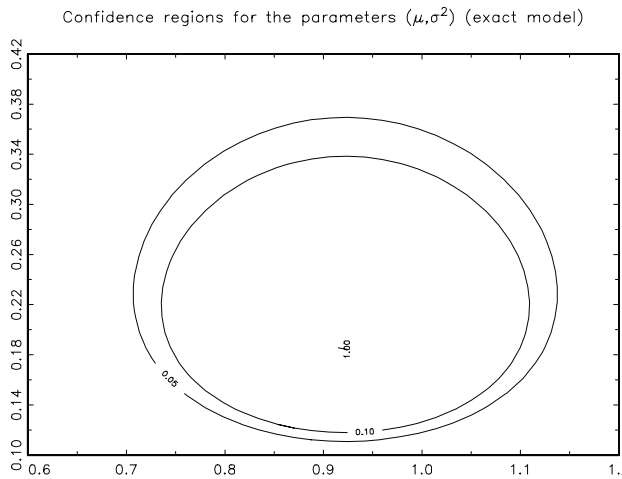

Figure 1.11. *Confidence estimates for* (μ, σ^2)

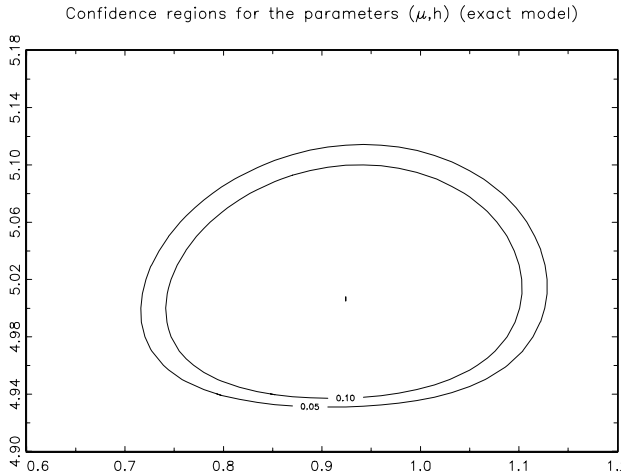

Figure 1.12. *Confidence estimates for* (μ, h)

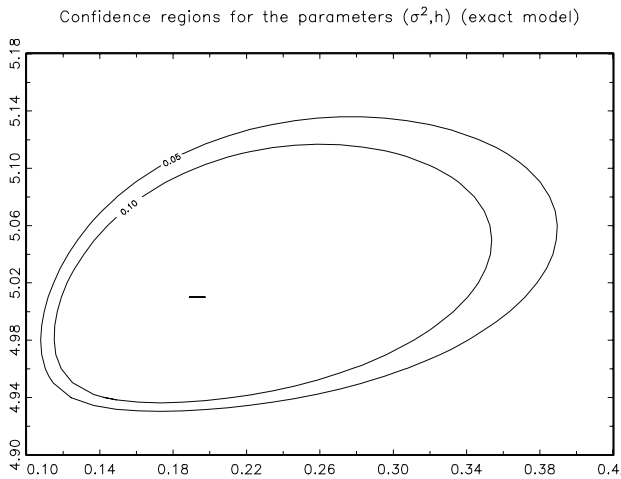

Figure 1.13. *Confidence estimates for* (σ^2, h)

Note that, in this case, it is also possible to estimate parameter h.

In an intensive further simulation study, we have estimated the partial parameter vector $\theta^* = (\mu, \sigma^2, h)$, whereas t_0 and x_0 were assumed to be known. Given an observation vector $X = (X_1, \ldots, X_n)$ with $X^{(i)} = (\min(\tau_i, t_{i,m_i}), x_{i,1}, \ldots, x_{i,M_i})$, only the maximum likelihood estimator for the drift parameter μ from the likelihood functions defined by equations [1.35] and [1.36] can be explicitly obtained as follows:

$$\widehat{\mu} = \frac{\displaystyle\sum_{i=1}^{n} \left(x_{i,M_i} - x_{i,0} + \mathbf{1}_{\{\tau_i < t_{i,m_i}\}} (\widehat{h} - x_{i,M_i}) \right)}{\displaystyle\sum_{i=1}^{n} \left(t_{i,M_i} - t_{i,0} + \mathbf{1}_{\{\tau_i < t_{i,m_i}\}} (\tau_i - t_{i,M_i}) \right)},$$

where $M_i = \max\{k : t_{i,k} < \tau_h^{(i)}\}$. The estimators $\widehat{\sigma^2}$ and $\widehat{h}$ were determined numerically.

Table 1.3 summarizes the estimated mean-squared errors of the maximum likelihood estimators $\widehat{\mu}$, $\widehat{\sigma^2}$ and $\widehat{h}$ from 1,000 simulation runs and compares them with those of the estimators $\widehat{\mu}_D$ and $\widehat{\sigma^2}_D$ from the likelihood function given by equation [1.4] that served as initial values for the optimization procedure. In each simulation run, $n = 50$ Wiener processes were simulated with $m_i = 10$ observation times $t_{i,j}$ and equal distances $t_{i,j} - t_{i,j-1} = 1$. The true parameter values were $\mu = 5$, $\sigma^2 \in \{1, 2, 5\}$, $t_0 = 0$, $x_0 = 0$ and $h \in \{20, 40, 50\}$. Additionally, Table 1.3 gives the average number $\bar{d}$ of observation times $t_{i,j}$ per degradation process before a failure.

Table 1.3 shows that, for all parameter choices in the simulation study, the mean-squared errors of the estimators $\widehat{\mu}$, $\widehat{\sigma^2}$ and $\widehat{h}$ based on the likelihood functions defined by equations [1.35] and [1.36] are smaller than those of the estimators $\widehat{\mu}_D$ and $\widehat{\sigma^2}_D$ based on the likelihood function in equation [1.4]. In particular, the general model based on the observation of degradation increments and failure times provides better estimates of the process parameters μ and σ^2, especially if nearly all items have failed during the experiment and if the number of actually observed degradation levels is small in comparison with the number of planned observation times.

μ	σ^2	h	$\bar{d}$	Empirical mean squared errors				
				$\widehat{\mu}$	$\widehat{\mu}_D$	$\widehat{\sigma^2}$	$\widehat{\sigma^2}_D$	$\widehat{h}$
		20	3.50	0.0107	0.0198	0.0347	0.0440	0.0028
5	2	40	7.55	0.0046	0.0062	0.0191	0.0196	0.0027
		50	9.34	0.0041	0.0042	0.0161	0.0172	0.0103
	1		3.51	0.0046	0.0085	0.0093	0.0114	0.0015
5	2	20	3.50	0.0107	0.0198	0.0347	0.0440	0.0028
	5		3.45	0.0358	0.0454	0.2992	0.5238	0.0043

Table 1.3. *Mean squared errors of maximum likelihood estimators in 1,000 simulation runs*

1.7. Extension and related models

Although it is widely used as a degradation model, the Wiener process suffers of some limitation and so it cannot be used anytime. One major drawback is that the average degradation is linear over time (see equation [1.3]). Some extensions or related models have hence been proposed in the literature, to remove this limitation. We here review two of them. With that aim, we introduce a time-change function $A : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ that is assumed to be increasing, continuous and such that $A(0) = 0$ and $\lim_{t\to\infty} A(t) = \infty$.

1.7.1. *Time-scaled Wiener processes*

A first extension consists of changing the time scale. More precisely, starting from a Wiener process $(Y_t)_{t\geq 0}$ with unit mean:

$$\forall t \geq 0, \quad Y_t = t + \sigma B_t,$$

a time-scaled Wiener process is defined as follows:

$$\forall t \geq 0, \quad X_t = Y_{A(t)} = A(t) + \sigma B_{A(t)}.$$

It is easy to check that, for any $t \geq 0$, X_t has the Gaussian distribution with mean $A(t)$ and variance $\sigma^2 A(t)$.

The non-overlapping increments of such a model are still independent but they are not stationary any more. Hence, the process remains additive (but it is no longer a Lévy process). Anyway, the likelihood function can easily be written based on the increments and maximized for a parametric form of A. A semi-parametric estimator has also been proposed by Wang [WAN 09b] based on a profile pseudo-log-likelihood approach (just as he did for the gamma process, see section 2.8.1 for details). Note that Wang has also considered the case of random effects for this model: in [WAN 09b], he has assumed that A is random and inverse Gaussian distributed, while in [WAN 10], he has considered a random drift and volatility with extra parameters to be estimated (hierarchical parameters).

The distribution of the time to failure can easily be expressed with respect to the inverse Gaussian distribution. Indeed, setting $\tau_{h,X}$ and $\tau_{h,Y}$ to be the time to failures associated with $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$, respectively, it is readily seen that $\tau_{X,h} = A^{-1}(\tau_{Y,h})$, where A^{-1} is the inverse function of A. Recalling that $\tau_{Y,h}$ has the inverse Gaussian distribution with parameters $(\beta = h^2/\sigma^2, \mu_1 = h)$ and using the expression of its p.d.f. given by equation [1.11], we have:

$$\forall t \geq 0, \quad f_{\tau_{X,h}}(t) = \frac{hA'(t)}{\sqrt{2\pi\sigma^2 A(t)^3}} \exp\left(-\frac{(h - A(t))^2}{2\sigma^2 A(t)}\right).$$

The mean time to failure may be quite difficult to compute explicitly, except for some particular cases (e.g. for a power transformation $A(t) = at^\gamma$).

1.7.2. *Brownian motion with nonlinear drift*

Instead of using a time-scaling, only the linear drift μt of the Wiener process can be replaced in equation [1.3] by the time-change $A(t)$. Hence, the model turns out to be:

$$\forall t \geq 0, \quad X_t = A(t) + \sigma B_t.$$

For any $t \geq 0$, X_t has the Gaussian distribution with mean $A(t)$, but with variance $\sigma^2 t$. Here again, the process remains additive, but it is not a Lévy process. The likelihood function can easily be written based on the independence of the increments. This model has been studied, for instance, by Si *et al.* [SI 12] (see also [WAN 14]).

Here, the distribution of the time to failure cannot be expressed explicitly for any choice of the function A. This can be viewed as the problem of determining the first passage time distribution of a time-dependent boundary by a Brownian motion (see [WAN 97], for instance). For any general boundary, some numerical approximations have been proposed in the literature. For instance, Wang and Pötzelberger [WAN 07] have proposed a computational approach based on a linear piecewise approximation for A and using Monte Carlo simulations.