Lecture 2 - Mathematical Foundations of Stochastic Processes

Charles R. Doering; notes by Gunnar Peng & Chris Spalding

June 16, 2015

1 Examples for last lecture

The moments do not determine the distribution. Consider the log-normal distribution and perturbed versions, whose PDFs are

$$f_0(x) = (2\pi)^{1/2}x^{-1}e^{-(\log x)^2/2}, \qquad f_a(x) = f_0(x)[1 + a\sin(2\pi\log x)] \text{ with } |a| \le 1. \tag{1}$$

It can be verified that these have the same moments – the nth moment is $e^{n^2/2}$ – even though the distributions are different. (Roughly speaking, the moments do determine the distribution if they do not grow too quickly, and evidently $e^{n^2/2}$ is too quick.)

Uncorrelation does not imply independence. The random variables

$$X \sim \text{Uniform}[-1, 1] \qquad \text{and} \qquad Y = X^2 \tag{2}$$

are clearly not independent. However, if we consider their mutual covariance,

$$\mathbb{E}[XY] = \mathbb{E}[X^3] = 0, \qquad \mathbb{E}[X] = 0, \qquad \mathbb{E}[Y] = \mathbb{E}[X^2] = 1/3 \qquad \Rightarrow \qquad \mathbb{E}[XY] = 0 = \mathbb{E}[X]\mathbb{E}[Y], \tag{3}$$

it becomes apparent that the variables are uncorrelated.

2 Brownian motion / Wiener process (continued)

Recall. The Wiener process $W(t) = W_t$ has transition probabilities and initial condition

$$\rho(w, t \mid w', t') = \frac{1}{\sqrt{2\pi(t-s)}}\exp\left(-\frac{1}{2}\frac{(w-w')^2}{t-s}\right) \text{ for } t \ge s, \qquad \rho(w, 0) = \delta(w), \tag{4}$$

which is illustrated in Figure 2. From this *continuous* distribution we can recover the *discretized*, n-time, probability density

$$\rho(w_n, t_n; \ldots; w_1, t_1) = \rho(w_n, t_n \mid w_{n-1}, t_{n-1}) \ldots \rho(w_2, t_2 \mid w_1, t_1)\rho(w_1, t_1) \qquad \text{for} \qquad t_n > \cdots > t_1. \tag{5}$$

Note that this is a natural factorization into independent increments, reflecting the Markovian property that each future increment is independent of the past.

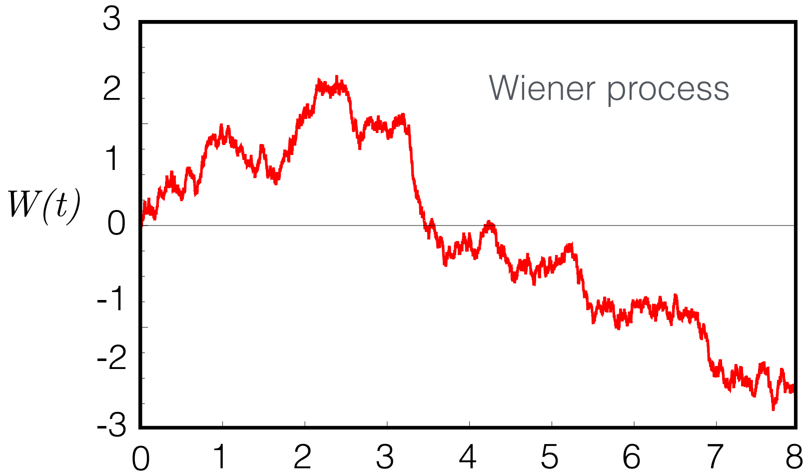

Figure 1: A typical realization of a Wiener process beginning at $W(0) = 0$. The step size used was $\Delta t = 0.001$.

To simulate the Wiener process on a computer, we discretize time from, say, $t_0 = 0$ to t_N, and start with $w_0 = 0$. Given the current value w_n, we repeatedly draw the next random value w_{n+1} using the probability distribution $\rho(w_{n+1}, t_{n+1} \mid w_n, t_n)$ given by Equation 4. This yields a single realization of the process; an example is shown in figure 1.

For comparison, Figure 3 displays both the continuous distribution (Equation 4) and a discrete distribution, obtained by sampling multiple realizations of the Wiener process (a typical example of which is shown in Figure 1). As is apparent, the continuous and discrete distributions look very similar, at least with the sample size of 10,000 used here. Note that the size of the time steps used to obtain realizations of the process does not affect the distribution, since the distribution obtained is given by equation 5, which is not an approximation.

Moments. The odd moments of a Wiener process vanish by symmetry, while the even moments may be calculated through integration by parts:

$$\mathbb{E}\left[W(t)^{2n+1}\right] = \int_{-\infty}^{\infty} w^{2n+1} \rho(w,t)\,\mathrm{d}w = 0, \tag{6}$$

$$\begin{aligned} \mathbb{E}\left[W(t)^{2n}\right] &= \int_{-\infty}^{\infty} w^{2n} \rho(w,t)\,\mathrm{d}w = \frac{(2n)!}{2^n n!} t^n \\ &= (2n-1)\cdot(2n-3)\cdot\dots\cdot 3\cdot 1\, t^n, \end{aligned} \tag{7}$$

$$\tag{8}$$

$$\text{e.g.} \qquad \mathbb{E}[W(t)^2] = t, \qquad \mathbb{E}[W(t)^4] = 3t^2, \qquad \mathbb{E}[W(t)^6] = 15t^3. \tag{9}$$

Correlation: It can be shown that, for $t \geq s$,

$$\mathbb{E}[W(t)W(s)] \equiv \int_{-\infty}^{\infty} \mathrm{d}w \int_{-\infty}^{\infty} \mathrm{d}w'\, w\, w'\, \rho(w,t;w',s) = s, \tag{10}$$

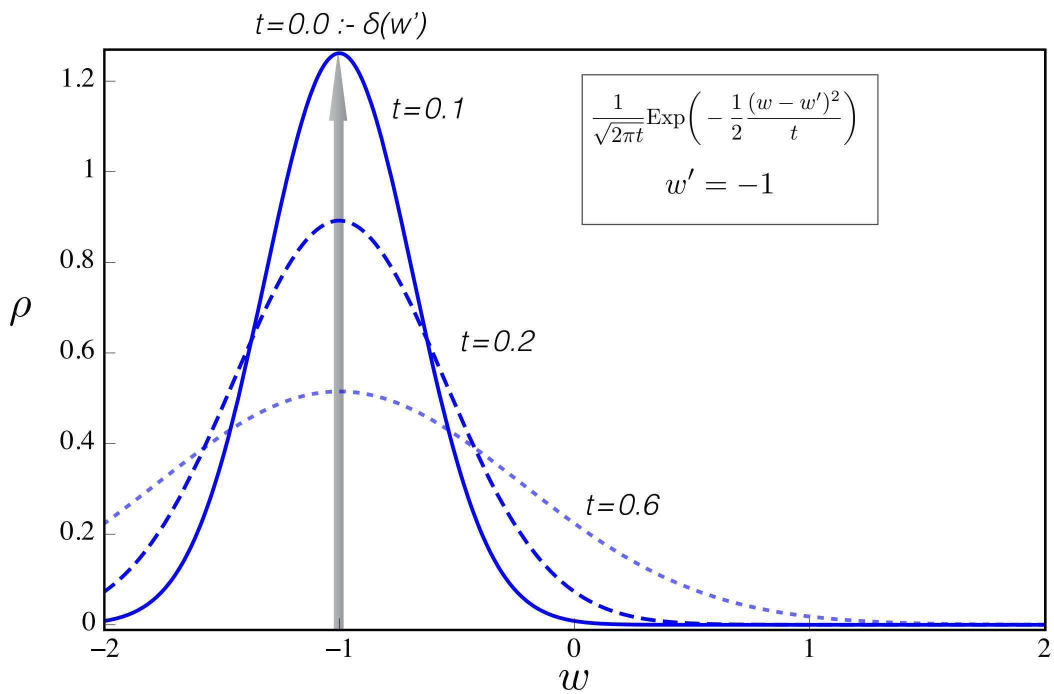

Figure 2: Illustration of the transition probability (density) appropriate for a Wiener process as a function of time. The initial distribution, set at time $t = s = 0$, is a delta function centered on $w' = -1$.

and hence, in general $\mathbb{E}[W(t)W(s)] = \min(t, s)$.

Exercise: Prove it the old-fashioned way (i.e., by changing variables and integrating)! One may get the answer through the slicker method below:

$$\begin{aligned}\mathbb{E}[W(t)W(s)] &= \mathbb{E}[\{W(t) - W(s)\}W(s) + W(s)^2] \\ &= \mathbb{E}[W(t) - W(s)]\mathbb{E}[W(s)] + s = s,\end{aligned} \tag{11}$$

where we used that the increment $W(t) - W(s)$ is independent of the past $W(s)$ provided that $t \geq s$.

2.1 Discrete analysis

Consider a discrete increment $\Delta W(t) = W(t + \Delta t) - W(t)$. It has mean and variance

$$\mathbb{E}[\Delta W(t)] = \underbrace{\mathbb{E}[W(t+\Delta t)]}_{0} - \underbrace{\mathbb{E}[W(t)]}_{0} = 0, \tag{12a}$$

$$\mathbb{E}[\Delta W(t)^2] = \underbrace{\mathbb{E}[W(t+\Delta t)^2]}_{t+\Delta t} - 2\underbrace{\mathbb{E}[W(t+\Delta t)W(t)]}_{t\,\text{(the earlier time)}} + \underbrace{\mathbb{E}[W(t)^2]}_{t} = \Delta t. \tag{12b}$$

This suggests that $\Delta W(t) = O(\sqrt{\Delta t})$. Hence $\Delta W/\Delta t = O(1/\sqrt{\Delta t})$ which will not converge as we take the continuum limit. Hence, $W(t)$ is not differentiable. However, there is no reason we can't simply tabulate the random variable $\Delta W(t)/\Delta t$ for many time-steps of finite width $\Delta t > 0$. We call this Gaussian white noise (GWN). From realizations of $W(t)$ (such as the one illustrated in Figure 1), we can explicitly calculate $\Delta W/\Delta t$ (see Figure 4), provided that Δt is a multiple of the time step used to generate $W(t)$. The GWN thus created remains confined to the 'x'-axis - reflecting its statistically steady nature. Deviations from zero occur with a typical magnitude that is of order $1/\sqrt{\Delta t}$ where Δt^{-1} is the rate at which the derivative is sampled.

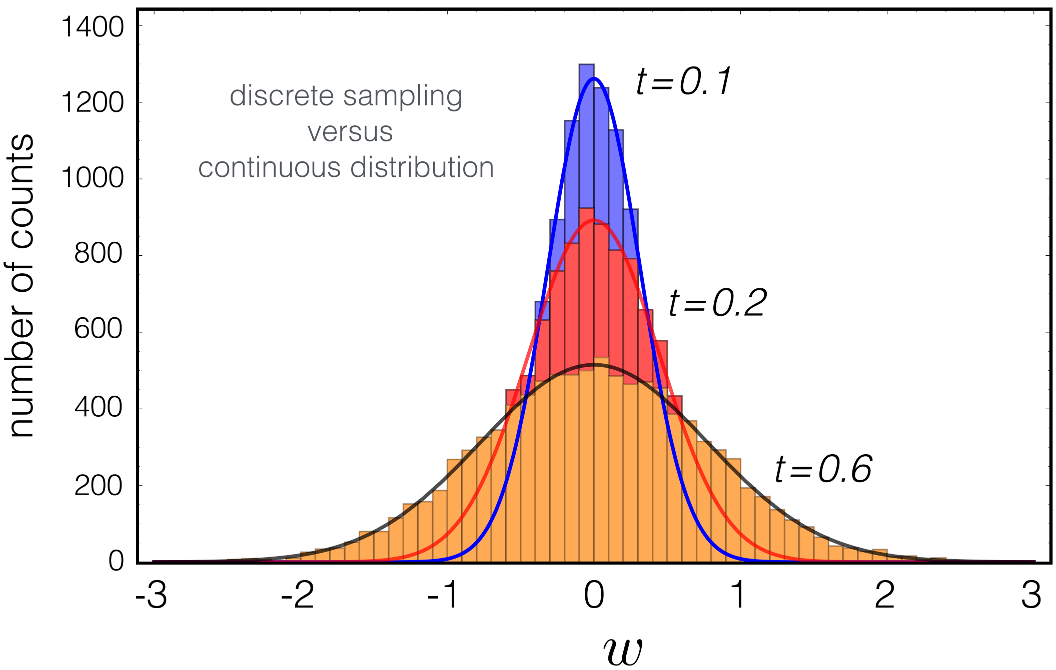

Figure 3: A comparison between the continuous distribution (as plotted in Figure 2) and the discrete distribution obtained by sampling multiple realizations of the Wiener Process (such as the one shown in Figure 1) starting with $W = 0$ at $s = 0$. Specifically, we recorded the value of $W(t)$ for $N = 10,000$ realizations of the Wiener process at the three values of time illustrated in Figure 2 ($t = 0.1, 0.2$ and 0.6). The resulting histograms, binned in groups of width 0.1, are presented. The respective continuous distributions, with mean $w' = 0$, are superimposed for comparison.

2.2 Continuous analysis

We (formally) define Gaussian white noise as

$$\xi(t) \equiv \frac{dW}{dt}. \tag{13}$$

This is a *distribution*-valued process, in contrast to a *function*-valued process. Hence, we investigate it through integration against smooth test functions (with compact support). If

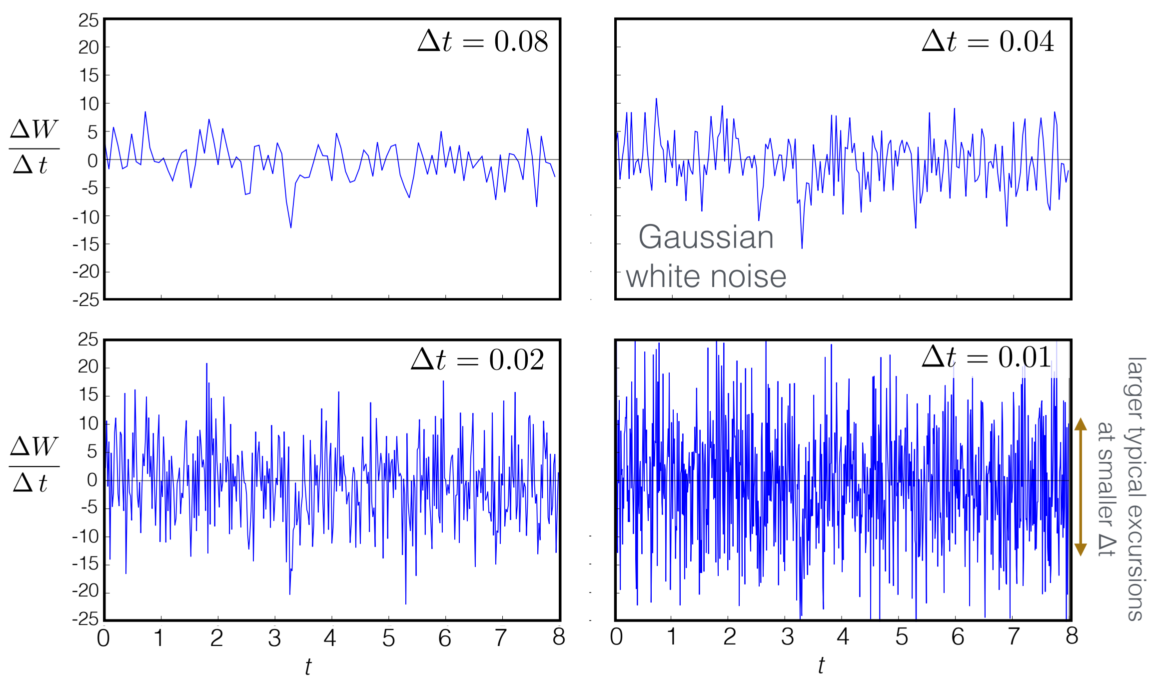

Figure 4: Numerical calculation of $\Delta W(t)/\Delta t$ sampled from the specific realization of $W(t)$ that is illustrated in Figure 1. We sample with 4 different values of the discrete time interval Δt (0.08, 0.04, 0.02 and 0.01). Gaussian white noise (GWN) is defined through such an operation. The GWN arising out of the Wiener process is statistically steady and confined to the 'x'-axis. Fluctuations from zero occur with a typical magnitude of the same order as $\sigma^{-1} = (\Delta t)^{-1/2}$. When compared to Figure 1 just after $t = 3$, for example, $W(t)$ begins to fall at the point where a large negative excursion occurs in $\Delta W/\Delta t$.

$\varphi(t)$ is such a function, then we define

$$\xi[\varphi] \equiv \int_0^\infty \xi(t)\varphi(t)\,\mathrm{d}t \equiv -\int_0^\infty W(t)\varphi'(t)\,\mathrm{d}t \qquad \text{(using integration by parts).} \tag{14}$$

This is a Gaussian random variable with mean $\mathbb{E}[\xi[\varphi]] = -\int \mathbb{E}[W(t)]\varphi'(t)\,\mathrm{d}t = 0$. (We assume that integrals and expectations commute when necessary.) For the variance, we compute

$$\mathbb{E}\left[\xi[\varphi]^2\right] = \mathbb{E}\left[\int_0^\infty W(t)\varphi'(t)\,\mathrm{d}t \int_0^\infty W(s)\varphi'(s)\,\mathrm{d}s\right] = \int_0^\infty\int_0^\infty \underbrace{\mathbb{E}\left[W(t)W(s)\right]}_{\min(t,s)}\varphi'(t)\varphi'(s)\,\mathrm{d}t\,\mathrm{d}s. \tag{15}$$

We simplify the s-integral first, by integrating by parts:

$$\int_0^\infty \varphi'(s)\min(t,s)\,\mathrm{d}s = -\int_0^t \varphi(s)\,\mathrm{d}s \qquad \Rightarrow \qquad \mathbb{E}\left[\xi[\varphi]^2\right] = -\int_0^\infty \varphi'(t)\left(\int_0^t \varphi(s)\,\mathrm{d}s\right)\mathrm{d}t, \tag{16}$$

where the first equality came from noticing that $\min(t,s)$ as a function of s has a gradient of unity for $s<t$ and zero elsewhere.

A second integration by parts yields the variance,

$$\boxed{\mathbb{E}\left[\xi[\varphi]^2\right]=\int_0^\infty \varphi(t)^2\,\mathrm{d}t=\int_0^\infty\int_0^\infty \varphi(t)\varphi(s)\delta(t-s)\,\mathrm{d}t\,\mathrm{d}s.} \tag{17}$$

Comparing the final expression with the formal calculation

$$\mathbb{E}\left[\xi[\varphi]^2\right]=\mathbb{E}\left[\int_0^\infty \xi(t)\varphi(t)\,\mathrm{d}t\int_0^\infty \xi(s)\varphi(s)\,\mathrm{d}s\right]=\int_0^\infty\int_0^\infty \mathbb{E}\left[\xi(t)\xi(s)\right]\varphi(t)\varphi(s)\,\mathrm{d}t\,\mathrm{d}s \tag{18}$$

suggests the formal result

$$\boxed{\mathbb{E}[\xi(t)\xi(s)]=\delta(t-s).} \tag{19}$$

A similar calculation yields $\mathbb{E}\left[\xi[\varphi]\xi[\psi]\right]=\int_0^\infty \varphi(t)\psi(t)\,\mathrm{d}t$.

We note that white noise is stationary, i.e. its statistics are independent of time. For any stationary process, the covariance is

$$C(t,s)\equiv\mathbb{E}[X(t)X(s)]-\mathbb{E}[X(t)]\mathbb{E}[X(t)]=\mathbb{E}[X(t-s)X(0)]-\mathbb{E}[X(0)]^2=c(t-s), \tag{20}$$

i.e. a function of $t-s$ only. We define the power spectrum as the Fourier transform

$$S(\omega)=\int_{-\infty}^{\infty} c(t)e^{-i\omega t}\,\mathrm{d}t. \tag{21}$$

For Gaussian white noise,

$$c(t-s)=\mathbb{E}[\xi(t)\xi(s)]=\delta(t-s)\qquad\Rightarrow\qquad S(\omega)=\int_{-\infty}^{\infty}\delta(t)e^{-i\omega t}\,\mathrm{d}t=1. \tag{22}$$

The noise is called “white” because it has a flat power spectrum, i.e. the same amount of energy at each frequency.

3 Stochastic differential equations, a.k.a. Langevin equations

A stochastic differential equation (SDE) is one that contains noise terms $\xi(t)$.

Example. What is the solution $X(t)$ to the equation

$$\frac{\mathrm{d}X}{\mathrm{d}t}=\xi(t), \tag{23}$$

with $X(0)=0$? It's clearly $X(t)=W(t)$!

Now, each $X(t)$ is but one random realization of the variable X described by the transition density ρ. In order to describe the spatial and temporal evolution of an ensemble of such realizations of $X(t)$, it is appropriate to ask how ρ itself evolves. For a pure Wiener process **(Exercise)**, show that the transition density satisfies the diffusion equation and the initial condition

$$\rho(x,t \mid y,s) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{1}{2}\frac{(x-y)^2}{t-s}\right) \qquad \Rightarrow \qquad \frac{\partial \rho}{\partial t} = \frac{1}{2}\frac{\partial^2 \rho}{\partial x^2} \quad \text{for} \quad t > s,$$
$$\rho(x,s) = \delta(x-y). \tag{24}$$

Since distributions can only be manipulated linearly, the most general SDE (without time delay) has the form

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = f(X(t),t) + g(X(t),t)\xi(t), \tag{25a}$$

which some mathematicians write (in a futile attempt to avoid differentiating non-differentiable things) as

$$\mathrm{d}X_t = f(X_t,t)\,\mathrm{d}t + g(X_t,t)\,\mathrm{d}W_t. \tag{25b}$$

The first term (with the function f) represents a deterministic, 'drift', term such as may be found in a regular dynamical systems equation. The second term describes stochastic noise with amplitude g.

The equivalent integral forms may be written as

$$X(t) = X(0) + \int_0^t f(X(s),s)\,\mathrm{d}s + \int_0^t g(X(s),s)\xi(s)\,\mathrm{d}s \tag{26a}$$
$$X_t = X_0 \quad + \int_0^t f(X_s,s)\,\mathrm{d}s \quad + \int_0^t g(X_s,s)\,\mathrm{d}W_s. \tag{26b}$$

(Note that the limits on the last integral indicate the range of s, not W_s.) The last integrals, involving the stochastic terms $\xi(s)$ or $\mathrm{d}W_s$, are ambiguous, and we will find that they may give different answers depending on how they are interpreted.

We mainly work with the Itō interpretation, in which (25) is defined as the continuous-time limit of the discrete-time system

$$\Delta X(t) \equiv X(t+\Delta t) - X(t) = f(X(t),t)\,\Delta t + g(X(t),t)\,\Delta W(t), \tag{27}$$

where $\Delta W(t) = W(t+\Delta t) - W(t) \sim N(0,\sqrt{\Delta t})$ (with $X(0)$ determined by an initial condition).

3.1 Conditional expectations

Recall. For $t \geq s$, the conditional expectation

$$\mathbb{E}\left[F(X(t)) \mid X(s) = y\right] = \int_{-\infty}^{\infty} F(x)\rho(x,t \mid y,s)\,\mathrm{d}x. \tag{28}$$

For example,

$$\mathbb{E}\left[W(t) \mid W(s) = w'\right] = \int_{-\infty}^{\infty} w\,\rho(w,t \mid w',s)\,\mathrm{d}w = w'. \tag{29}$$

We now compute the expectation $\mathbb{E}\left[\Delta X(t) \;\middle|\; X(t)=x\right]$ in two different ways. Firstly, in terms of the density:

$$\mathbb{E}\left[\Delta X(t) \;\middle|\; X(t)=x\right] = \mathbb{E}\left[X(t+\Delta t) \;\middle|\; X(t)=x\right] - \underbrace{\mathbb{E}\left[X(t) \;\middle|\; X(t)=x\right]}_{x} = \tag{30a}$$

$$= \int x'\, \rho(x', t+\Delta t \mid x, t)\,\mathrm{d}x' - x \underbrace{\int \rho(x', t+\Delta t \mid x, t)\,\mathrm{d}x'}_{\text{1 in disguise}} = \tag{30b}$$

$$= \int (x'-x)\rho(x', t+\Delta t \mid x, t)\,\mathrm{d}x'. \tag{30c}$$

Secondly, making use of (27):

$$\mathbb{E}\left[\Delta X(t) \;\middle|\; X(t)=x\right] = \mathbb{E}\left[f(X(t),t) \;\middle|\; X(t)=x\right]\,\Delta t + \mathbb{E}\Bigg[\underbrace{\underbrace{g(X(t),t)}\underbrace{\Delta W(t)}}_{\text{independent}} \;\Bigg|\; X(t)=x\Bigg] = \tag{31a}$$

$$= f(x,t)\,\Delta t + \underbrace{\mathbb{E}\left[g(X(t),t) \;\middle|\; X(t)=x\right]}_{g(x,t)} \underbrace{\mathbb{E}\left[\Delta W(t) \;\middle|\; X(t)=x\right]}_{0} = f(x,t)\,\Delta t \tag{31b}$$

Hence, we conclude that

$$\boxed{\int_{-\infty}^{\infty} (x'-x)\,\rho(x', t+\Delta t \mid x, t)\,\mathrm{d}x' = f(x,t)\,\Delta t.} \tag{32}$$

Similarly, we compute the expectation $\mathbb{E}\left[\Delta X(t)^2 \;\middle|\; X(t)=x\right]$ in two different ways:

$$\mathbb{E}\left[\Delta X(t)^2 \;\middle|\; X(t)=x\right] = \int (x'-x)^2 \rho(x', t+\Delta t \mid x, t)\,\mathrm{d}x' \quad \text{as before, and} \tag{33a}$$

$$\mathbb{E}\left[\Delta X(t)^2 \;\middle|\; X(t)=x\right] = \mathbb{E}_W \int \left[f(x',t)\,\Delta t + g(x',t)\,\Delta W(t)\right]^2 \rho(x', t+\Delta t | x, t)\,\mathrm{d}x' = \tag{33b}$$

$$= f(x,t)^2\,\Delta t^2 + g(x,t)\underbrace{\mathbb{E}_W[\Delta W(t)]}_{0} + g(x,t)^2 \underbrace{\mathbb{E}_W[\Delta W(t)^2]}_{\Delta t}. \tag{33c}$$

Hence,

$$\boxed{\int_{-\infty}^{\infty} (x'-x)^2\,\rho(x', t+\Delta t \mid x, t)\,\mathrm{d}x' = f(x,t)^2\,\Delta t^2 + g(x,t)^2\,\Delta t.} \tag{34}$$