Appendix A

Exponentials of Gaussian random variables

Gauge P simulations of interacting Bose gases (whether dynamic or thermodynamic) involve multiplicative noise terms of the form

$$dz = czdW(t) + \dots, \tag{A.1}$$

where z is a complex variable ($\alpha_{\mathbf{n}}$, $\beta_{\mathbf{n}}$, z_0), and c a constant. This leads to $\log z$ being distributed approximately as a Gaussian. This is exact at short times, but some modification of the distribution occur also as a consequence of the other "…" drift (or noise) terms in the equations. Observable estimates are usually obtained through averages of non-logarithmic combinations of z, however, and this requires some care. Let us consider the idealized situation of $z \approx v_\sigma = v_0 e^{\sigma\xi} = e^{v_L}$, where ξ is a Gaussian random variable of mean zero, variance unity. The notation of Section 7.4 will be used.

Using the distribution of ξ, (7.36), one obtains

$$\Pr(v_\sigma) = \frac{1}{\sigma v_\sigma \sqrt{2\pi}} \exp\left\{ -\frac{(\log[v_\sigma/v_0])^2}{2\sigma^2} \right\}. \tag{A.2}$$

This distribution falls off slowly as $v_\sigma \to \infty$, and for a finite sample may not be sampled correctly if ξ is chosen according to its Gaussian distribution. With $\mathcal{S}$ samples, the greatest value of ξ obtained can be expected to be $\xi_{\max}$, the solution of $\frac{1}{2}[1 - \text{erf}(\xi_{\max}/\sqrt{2})] \approx 2/\mathcal{S}$, where $\text{erf}(x) = 2\int_0^x e^{-t^2} dt$ is the error function. For

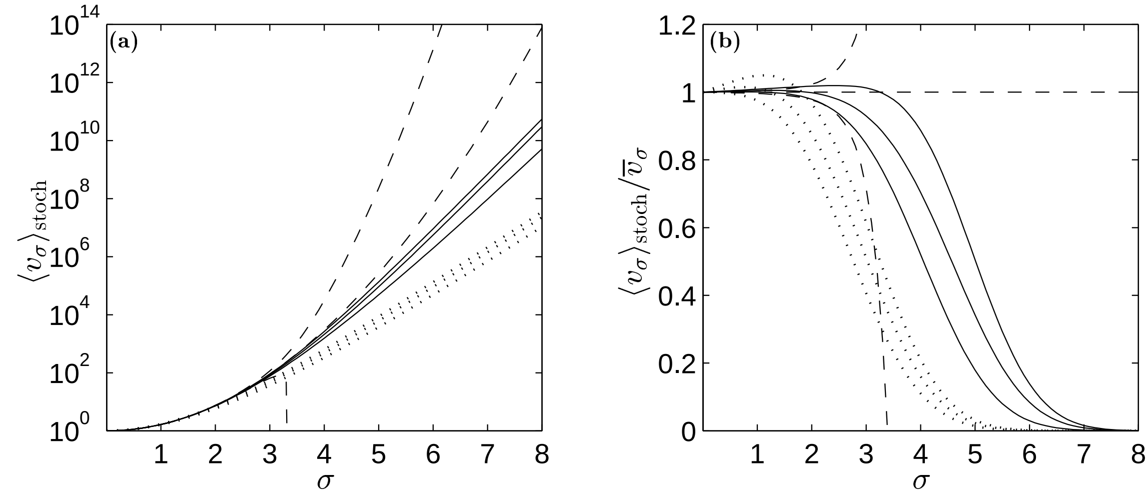

Figure A.1: **Finite sample estimates** of means of exponentials of gaussian random variables $v_\sigma = e^{\sigma\xi}$. SOLID lines: $\mathcal{S} = 10^5$ sample estimate of $\langle v_\sigma \rangle_{\rm stoch}$ and error bars, where ξ was sampled from a Gaussian distribution. DOTTED lines: Same but using $\mathcal{S} = 1000$ samples. DASHED lines: Exact ($\mathcal{S} \to \infty$) value of the mean $\langle v_\sigma \rangle \to \overline{v}_\sigma$, along with error estimates for the $\mathcal{S} = 10^5$ calculation based on the exact value of var $[v_\sigma]$, from (7.39) and (7.40). This corresponds to what would be expected from an unbiased sample of v_σ. All error bars were calculated according to (7.41). Subplot **(b)** shows the same values, but scaled with respect to the $\mathcal{S} \to \infty$ mean $\overline{v}_\sigma$. Values at all σ were made with the same sample of ξ to show the direct dependence on σ only.

$\mathcal{S} \approx \mathcal{O}(1000)$ to $\mathcal{O}(10^5)$, this gives $\xi_{\max} \approx 3$ to 4. Some bias in $\langle v_\sigma \rangle_{\rm stoch}$ (underestimation) may be expected due to the lack of sampling of the far tails once

$$\int_{v_0 \exp(\sigma\xi_{\max})}^{\infty} \Pr(v_\sigma)\, dv_\sigma \gg 1/\mathcal{S}. \tag{A.3}$$

For typical sample sizes $\mathcal{S} \approx \mathcal{O}(1000)$ to $\mathcal{O}(10^5)$ this will usually occur around $\sigma^2 \gtrsim 10$. This is actually the same large σ region given by (7.43) where all precision is lost anyway because of excessive distribution spread (as shown in Section 7.4).

However, the problem is that *both* the mean $\langle v_\sigma \rangle_{\rm stoch}$ and the CLT precision estimate $\sqrt{\mathrm{var}[v_\sigma]/\mathcal{S}}$ are underestimated when σ is large, but the error estimate is underestimated *by an even greater amount* than the mean. As a result, badly sampled data may still appear to be significant for large σ values. The situation is shown in Figure A.1.

The simplest practical solution is to simply discard any calculated means $\langle z \rangle_{\rm stoch}$

when

$$\mathrm{var}\left[\log|z|\right] \gtrsim 10. \tag{A.4}$$

If an unbiased mean $\langle z\rangle_{\mathrm{stoch}}$ and precision estimate Δz was obtained by some more sophisticated method, then it would not be significant anyway since one expects $\Delta z \gg \langle z\rangle_{\mathrm{stoch}}$ at these large σ values. The main point here is that quantities (A.4) should be monitored when dealing with multiplicative noise of the form (A.1) in the stochastic equations.

Some subtleties can arise because in realistic simulations $\log z$ is not exactly Gaussian. If the large $\log z$ distribution tails fall off more rapidly than Gaussian, then an unbiased simulation can be obtained for larger $\mathrm{var}\left[\log|z|\right]$, while the converse is true if these tails fall of less rapidly. A **more robust** indicator of possible bias than (A.4) is to compare the observable estimate obtained with two sample sizes $\mathcal{S}$, differing by at least an order of magnitude. If sampling bias is present, the two estimates of the average $\langle z\rangle_{\mathrm{stoch}}$ will usually differ by a statistically significant amount. An example of this for the idealized v_σ model is shown in Figure A.1. This kind of sample-size-dependent behavior is always a strong warning sign.