The forward diffusion process is defined by:

$$q(x_t \mid x_{t-1}) = \mathcal{N}(x_t; \sqrt{\alpha_t} x_{t-1}, (1 - \alpha_t)\mathbf{I}) \tag{1}$$

where $\alpha_t \in (0, 1)$ controls the amount of signal retained at each step, and $(1-\alpha_t)$ determines the variance of the added Gaussian noise. Define $\beta_t = 1 - \alpha_t$ for convenience.

Recursive Substitution

By repeatedly applying the forward process from x_0 to x_t, we can marginalize out all intermediate steps. **The key insight from the DDPM paper is that the marginal distribution of x_t given x_0 is also Gaussian:**

$$q(x_t \mid x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t} x_0, (1 - \bar{\alpha}_t)\mathbf{I}) \tag{2}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$ (basically we run all the steps from 1 to t, which has a Gaussian marginal distribution).

Since the marginal distribution is known, we can sample x_t directly from x_0 using:

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t}\, \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}) \tag{3}$$

This is the equation used during training to generate noisy images x_t from clean images x_0 in a single step, making the process efficient.