Sampling form distributions

Let $f\colon \Omega \to \mathbb{R}^+$ be a probability density function on a domain $\Omega \subset \mathbb{R}$ and F be the corresponding cumulative distribution function

$$F(x) = \int_{-\infty}^{x} f(x)\, dx.$$

Sampling from f means producing a random variable X such that

$$\Pr(X < x) = F(x). \tag{1}$$

In general, this might not be easy to do. However, for specific distributions efficient algorithms might exist. In particular, practically every programming language has facilities for sampling from Uniform $[0, 1]$, which has pdf

$$f_{\text{uniform}}(y) = \begin{cases} 1 & \text{if } \ 0 \leq y < 1 \\ 0 & \text{otherwise} \end{cases}$$

and cdf

$$F_{\text{uniform}}(x) = \begin{cases} 0 & y < 0 \\ y & 0 \geq y < 1 \\ 1 & y \geq 1. \end{cases}$$

Assuming that Y is sampled from Uniform $[0, 1]$ then, $\Pr(Y < y) = y$ and for any strictly monotonic increasing function $g : [0, 1] \to \mathbb{R}$

$$\Pr(g(Y) < g(y)) = y.$$

Now note that $g(Y)$ is itself a random variable that we can call Z. Letting $z = g(y)$

$$\Pr(Z < z) = g^{-1}(z) \tag{2}$$

where g^{-1} is the inverse of g in the sense that $g^{-1}(g(y)) = y$. Equation (2) shows that the cdf of Z is just g^{-1}. Coming back to our original problem and letting $g = F^{-1}$ we see that $X = Z$ will be exactly the random variable satisfying (1) that we were looking for.

In the specific case of the normal distribution,

$$F(x) = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\bigl(x/\sqrt{2}\bigr)$$

where the error function $\operatorname{erf}(u)$ is defined

$$\operatorname{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-t^2} dt,$$

so $2F(x) + 1 = \operatorname{erf}\left(x/\sqrt{2}\right)$ and $\operatorname{erf}^{-1}(2y + 1) = F^{-1}(y)/\sqrt{2}$. Hence if $Y \sim$ Uniform $[0, 1]$ then $\sqrt{2}\operatorname{erf}^{-1}(2Y + 1) \sim \text{Normal}(0, 1)$.

Note that the the error function (or its inverse) cannot be expressed in closed form, so this is actually probably not the best way to sample from a Gaussian in practice. Instead we can just take N independent symmetric binary random variables $\tau \in \{-1, 1\}$. As N becomes large, by the law of large numbers the distribution of

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tau_i$$

will quickly tend to $\mathrm{Normal}(0, 1)$.