Random variables and random number generation

3F3 laboratory experiment
Theo A. Brown
Selwyn College, University of Cambridge

November 5, 2021

Abstract

Enter a short summary here.

1 Generating random numbers from the Uniform and Normal distributions

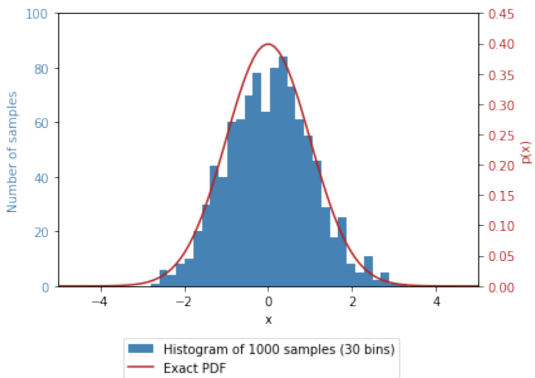
1.1 Comparison of histogram with true probability density function

A vector of Gaussian random numbers was generated using `np.random.randn`, and a vector of uniformly distributed random numbers generated using `np.random.rand`. Histograms of the samples are plotted in Figure 1, overlaid with the exact probability density function (PDF). The histograms closely follow the shape of the PDF, showing that the generation of the random numbers for these distributions is accurate.

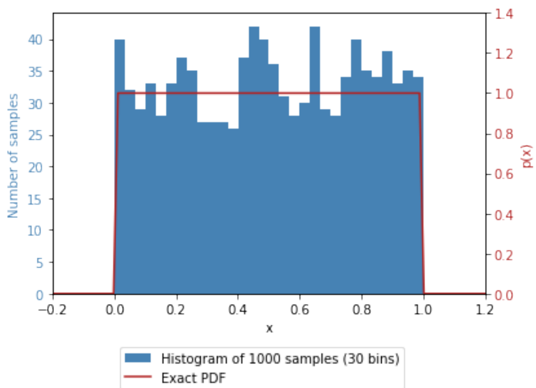
1.2 Kernel density smoothing

A smooth estimate of the probability density function is calculated using the kernel smoothing method with a Gaussian kernel $\mathcal{N}(0, 1)$. The result is plotted for Gaussian and Uniform distributions in Figure 3.

The kernel method takes an average of neighbouring values, weighted using a Gaussian distribution. This is advantageous as it can smooth random irregularities in the samples: there may be histogram bins with a disproportionately high sample count, which would give an incorrectly high estimate of the probability of this bin; the kernel smooths out this local peak by taking neighbouring values into account. However, this also means that the kernel smoothing method becomes inaccurate when there is a sudden change or discontinuity in the probability density function. In the Uniform distribution there is zero probability for values outside the range $0 < x < 1$, but as the kernel averages over a window of values it does not capture the step change at $x = 0$ and $x = 1$ and instead decays smoothly.

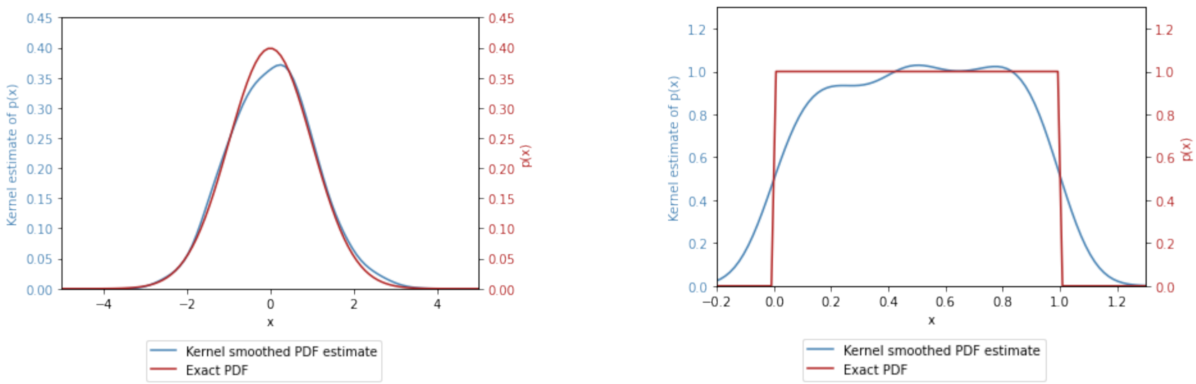


(a) Gaussian distribution



(b) Uniform distribution

Figure 1: Histogram of samples drawn from a distribution and true probability density function of distribution



(a) Gaussian distribution

(b) Uniform distribution

Figure 2: Plot of true PDF and estimate of PDF generated using Gaussian kernel smoothing

1.3 Multinomial distribution theory

For N samples, let B be the random variable representing the number of samples in bin j. Let the probability that a sample falls in bin j be p_j. By definition, the expected number of samples in bin j $\mu_j = \mathbb{E}[B] = Np_j$, and the variance in the number of samples in bin j $\sigma_j^2 = Var[B] = Np_j(1 - p_j)$.

For the Uniform distribution between 0 and 1, the pdf $p(x)$ is defined as:

$$
\begin{aligned}
p(x) &= \frac{1}{x_{max} - x_{min}} \\
&= 1
\end{aligned}
$$

Hence:

$$
\begin{aligned}
p_j &= \int_{c_j - \delta/2}^{c_j + \delta/2} dx \\
&= \delta \\
\mu_j &= N\delta \\
\sigma_j &= \sqrt{N\delta(1 - \delta)}
\end{aligned}
$$

In the histograms plotted in **??**,

2 Conclusions