Stochastic Numerical Analysis

Prof. Mike Giles

mike.giles@maths.ox.ac.uk

Oxford University Mathematical Institute

Probability basics

Defn: a scalar random variable X has continuous probability density $p(x)$ if

$$\mathbb{P}[X \in (x, x+\mathrm{d}x)] = p(x)\,\mathrm{d}x + o(\mathrm{d}x)$$

so we then have

$$\mathbb{E}[f(X)] = \int f(x)\,p(x)\,\mathrm{d}x.$$

Defn: the variance $\mathbb{V}[X]$ is

$$\mathbb{V}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

Probability basics

For any scalar random variable X and constants μ, λ,

$$
\begin{aligned}
\mathbb{E}[X+\mu] &= \mathbb{E}[X]+\mu \\
\mathbb{V}[X+\mu] &= \mathbb{V}[X] \\
\mathbb{E}[\lambda\, X] &= \lambda\, \mathbb{E}[X] \\
\mathbb{V}[\lambda\, X] &= \lambda^2\, \mathbb{V}[X]
\end{aligned}
$$

Probability basics

Defn: a pair of scalar random variables X, Y has joint p.d.f. $p(x,y)$ if

$$\mathbb{P}[X \in (x, x{+}\mathrm{d}x), Y \in (y, y{+}\mathrm{d}y)] = p(x,y)\ \mathrm{d}x\,\mathrm{d}y + o(\mathrm{d}x\,\mathrm{d}y)$$

and therefore

$$\mathbb{E}[f(X,Y)] = \int f(x,y)\ p(x,y)\ \mathrm{d}x\,\mathrm{d}y$$

We clearly have

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

Probability basics

We also have

$$\mathbb{V}[X+Y] = \mathbb{V}[X] + 2\,\text{Cov}[X,Y] + \mathbb{V}[Y]$$

where the covariance is defined as

$$\text{Cov}[X,Y] \equiv \mathbb{E}\Big[\,(X-\mathbb{E}[X])\,(Y-\mathbb{E}[Y])\Big]$$

Note: $\mathbb{E}[(X-\lambda Y)^2] \geq 0$, so put $\lambda = \sqrt{\mathbb{E}[X^2]/\mathbb{E}[Y^2]}$ and re-arrange to get

$$\mathbb{E}[X\,Y] \leq \sqrt{\mathbb{E}[X^2]\ \mathbb{E}[Y^2]}$$

and hence

$$\Big|\text{Cov}[X,Y]\Big| \leq \sqrt{\mathbb{V}[X]\ \mathbb{V}[Y]}$$

Probability basics

X, Y are independent iff the joint p.d.f. is a product of their individual p.d.f.'s:

$$p(x, y) = p_X(x)\; p_Y(y)$$

In this case we get

$$\mathbb{E}[f(X)\; g(Y)] = \mathbb{E}[f(X)]\; \mathbb{E}[g(Y)]$$

and then Cov$[X, Y] = 0$ and hence

$$\mathbb{V}[X+Y] = \mathbb{V}[X] + \mathbb{V}[Y]$$

Probability basics

Notation: $X \sim N(\mu, \sigma^2)$ means X has the distribution of a Normal random variable with mean μ and variance σ^2.

The p.d.f. is

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \; \exp\left(-\,\frac{(x-\mu)^2}{2\sigma^2}\right)$$

If

$$X \sim N(\mu, \sigma^2)$$

then

$$\begin{aligned} X + \lambda \;\; &\sim \;\; N(\mu+\lambda,\; \sigma^2) \\ \lambda X \;\; &\sim \;\; N(\lambda\mu,\; \lambda^2\sigma^2) \end{aligned}$$

Probability basics

If X, Y are independent, and $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$, then

$$X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

Proof: see Wikipedia article

Probability basics

If $Z \sim N(0,1)$ (a unit / standard Normal random variable) then for integer n we have

$$\mathbb{E}[Z^{2n+1}] = 0$$

and

$$\mathbb{E}[Z^{2n}] = \frac{(2n-1)!}{(n-1)!\, 2^{n-1}}$$

Proof: integration by parts

Monte Carlo estimation

Given a random variable X we want to estimate $\mathbb{E}[X]$.

We do so by taking N independent samples to obtain the estimate

$$E_N = N^{-1} \sum_{n=1}^{N} X^{(n)}.$$

Note that $\mathbb{E}[E_N] = \mathbb{E}[X]$ so it is an unbiased estimate.

Also, due to independence

$$\mathbb{V}[E_N] = N^{-1} \mathbb{V}[X]$$

Central Limit Theorem

Define

- error $\quad \varepsilon_N = E_N - \mathbb{E}[X]$
- RMSE, "root-mean-square-error" $= \sqrt{\mathbb{E}[\varepsilon_N^2]}$

Loosely speaking, the Central Limit Theorem proves that for large N

$$\varepsilon_N \sim \sigma\, N^{-1/2}\, Z$$

with Z a $N(0,1)$ random variable and $\sigma^2 = \mathbb{V}[X]$

Central Limit Theorem

More precisely, provided σ is finite, then as $N \longrightarrow \infty$,

$$\text{CDF}(N^{1/2}\sigma^{-1}\varepsilon_N) \longrightarrow \text{CDF}(Z)$$

so that

$$\mathbb{P}\left[N^{1/2}\sigma^{-1}\varepsilon_N < s\right] \;\longrightarrow\; \mathbb{P}\left[Z < s\right] \;=\; \Phi(s)$$

and

$$\mathbb{P}\left[\left|N^{1/2}\sigma^{-1}\varepsilon_N\right| > s\right] \;\longrightarrow\; \mathbb{P}\left[|Z| > s\right] \;=\; 2\,\Phi(-s)$$

$$\mathbb{P}\left[\left|N^{1/2}\sigma^{-1}\varepsilon_N\right| < s\right] \;\longrightarrow\; \mathbb{P}\left[|Z| < s\right] \;=\; 1 - 2\,\Phi(-s)$$

Variance estimation

Given N samples, the empirical variance is

$$\widetilde{\sigma}^2 = N^{-1} \sum_{n=1}^{N} \left(X^{(n)} - E_N \right)^2 = E_N^{(2)} - (E_N)^2$$

where

$$E_N = N^{-1} \sum_{n=1}^{N} X^{(n)}, \qquad E_N^{(2)} = N^{-1} \sum_{n=1}^{N} \left(X^{(n)} \right)^2$$

$\widetilde{\sigma}^2$ is a slightly biased estimator for σ^2; an unbiased estimator is

$$\widehat{\sigma}^2 = (N-1)^{-1} \sum_{n=1}^{N} \left(X^{(n)} - E_N \right)^2 = \frac{N}{N-1} \left(E_N^{(2)} - (E_N)^2 \right)$$

Confidence interval

Objective: want an accuracy of $\overline{\varepsilon}$ with confidence c.
i.e. $|\varepsilon| < \overline{\varepsilon}$ with probability c.

How many samples do we need to use?

Recall,

$$\mathbb{P}\left[N^{1/2}\sigma^{-1}|\varepsilon| < s \right] \approx 1 - 2\,\Phi(-s),$$

so define function $s(c)$ such that

$$1 - 2\,\Phi(-s) = c \quad \Longleftrightarrow \quad s = -\Phi^{-1}((1{-}c)/2)$$

Confidence interval

c	0.683	0.9545	0.9973	0.99994
s	1.0	2.0	3.0	4.0

Then $|\varepsilon| < N^{-1/2}\,\sigma\, s(c)$ with probability c, so to get $|\varepsilon| < \overline{\varepsilon}$ we can put

$$N^{-1/2}\,\widehat{\sigma}\, s(c) = \overline{\varepsilon} \quad \Longrightarrow \quad N = \left(\frac{\widehat{\sigma}\, s(c)}{\overline{\varepsilon}}\right)^2.$$

Note: twice as much accuracy requires 4 times as many samples.

Biased estimation

Sometimes unable to generate samples $X^{(n)}$ from the correct distribution, and instead generate samples $\widehat{X}^{(n)}$ from a similar distribution.

Estimator is then

$$E_N = N^{-1} \sum_{n=1}^{N} \widehat{X}^{(n)}$$

with expected value $\mathbb{E}[\widehat{X}]$ and variance $N^{-1}\mathbb{V}[\widehat{X}]$.

Biased estimation

The Mean Square Error is

$$
\begin{aligned}
\mathbb{E}\left[(E_N - \mathbb{E}[X])^2\right] &= \mathbb{E}\left[\left(E_N - \mathbb{E}[\widehat{X}] \;+\; \mathbb{E}[\widehat{X}] - \mathbb{E}[X]\right)^2\right] \\
&= \mathbb{E}\left[(E_N - \mathbb{E}[\widehat{X}])^2\right] + \left(\mathbb{E}[\widehat{X}] - \mathbb{E}[X]\right)^2 \\
&= N^{-1}\mathbb{V}[\widehat{X}] + \left(\mathbb{E}[\widehat{X}] - \mathbb{E}[X]\right)^2
\end{aligned}
$$

- first term is due to the variance of estimator
- second term is square of bias due to (weak) approximation error