数值积分大作业

蹇傲霖 2018010919 电 82

用数值积分方法计算标准正态分布分布函数 $\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$ 在一些节点上近似值，建立一个表格。将此表计算值相邻的两点用线段连接，用以估计对任意 $x\in R$，$\Phi(x)$的近似值。将你的 $\Phi(x)$近似值与 Matlab 中 normcdf 函数的计算结果进行对比。

- **关于广义积分**

对于

$$\int_{-\infty}^{m}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}t^2}\mathrm{d}t \qquad m<-2$$

可以用上界约束：

$$\int_{-\infty}^{m}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}t^2}\mathrm{d}t<\int_{-\infty}^{m}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{t}\mathrm{d}t=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{m}$$

令$\frac{1}{\sqrt{2\pi}}\mathrm{e}^{m}=10^{-6}$，解得 m=-12.8966，可取 m=-13，以保证从 m 开始的积分与真实值误差小于 10^{-6}。

- **认识目标函数**

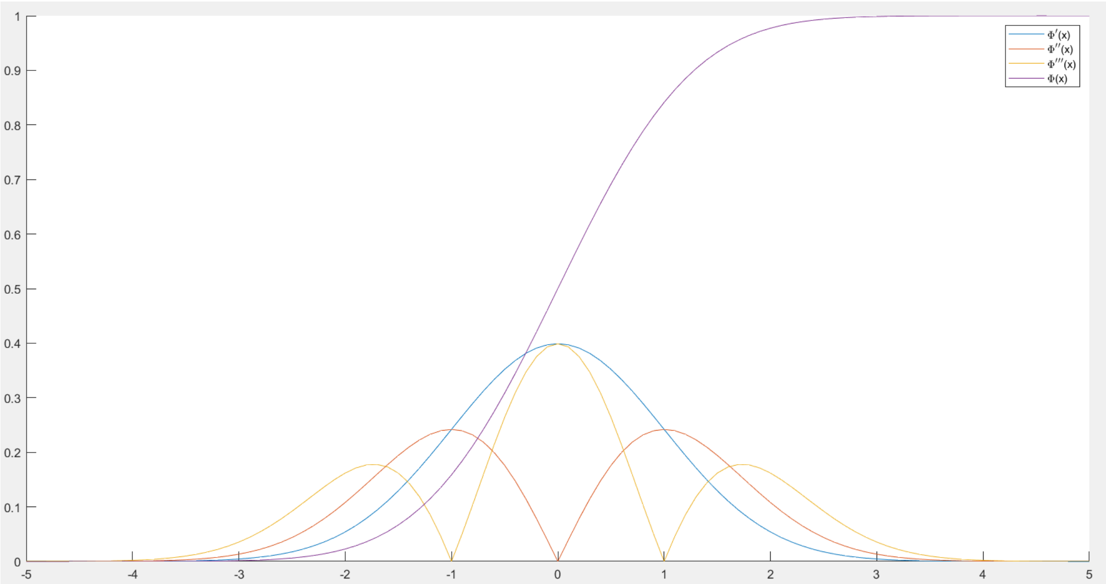
首先画出正态分布函数的 CDF、正态分布函数、正态分布函数的一阶导数、正态分布函数的二阶导数，一共四条曲线如下图所示，其中有负值的曲线取了绝对值。其中正态分布函数的 CDF 即$\Phi(x)$是由 MATLAB 函数 normcdf 所得到的。

另外三条曲线有解析式：

$$\Phi'(x)=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}$$

$$\Phi''(x)=\frac{-1}{\sqrt{2\pi}}x\mathrm{e}^{-\frac{1}{2}x^2}$$

$$\Phi'''(x)=\frac{-1}{\sqrt{2\pi}}(1-x^2)\mathrm{e}^{-\frac{1}{2}x^2}$$



我们要求解的就是$\Phi(x)$（紫色曲线），可以看出其满足

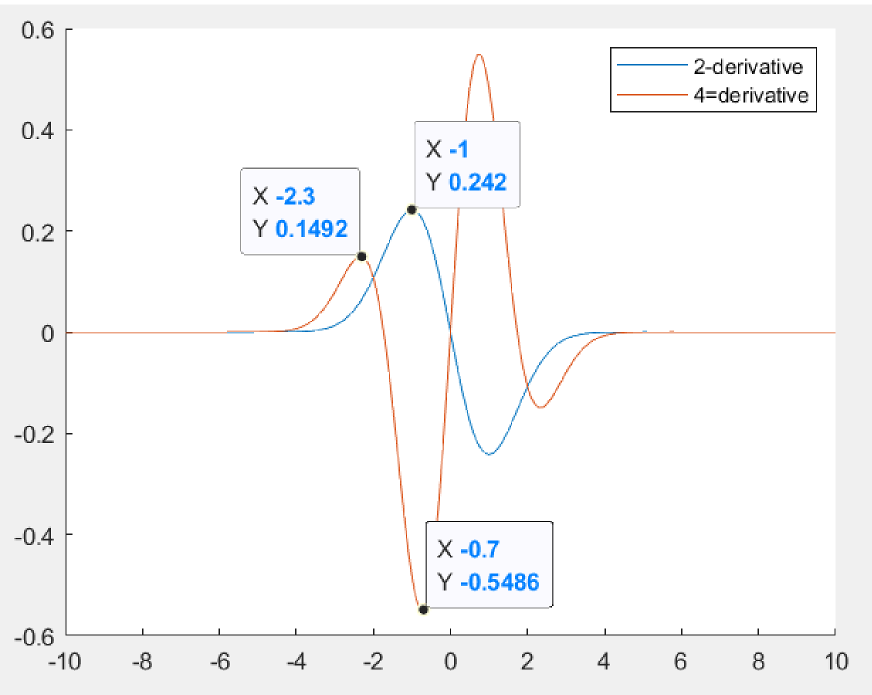
$$\Phi(x) + \Phi(-x) = 1$$

那么只需要在$(-\infty, 0]$上进行数值积分，进行翻转、平移、去零操作即可补全。

另外，可以明显地看出，在目标函数$\Phi(x)$（紫色）弯曲越严重的地方，对应$\Phi''(x)$（红色）的值越大，对应线性插值的误差也会越大。

- **关于二阶导数与四阶导数绝对值的最大值求解**

下图定性地画出了二阶导数和四阶导数的函数图像：



利用函数区间内最小值求解的 fminbnd 函数，在给定区间[-1.1,-0.9],[0.6,0.9]内求得精确最值：

$$\mathrm{M}_2 = \max|\Phi''(x)| = 0.241970724519047$$

$$\mathrm{M}_4 = \max|\Phi''''(x)| = 0.550587839498791$$

且极值点对应为: -1 ; -0.742 , -2.3344

- **关于线性插值**

上面函数图像弯曲程度与线性插值误差之间关系的定性认识，用以下引理可以进行定量说明：

用线段连接$\left(a, f(a)\right)$和$\left(b, f(b)\right)$两点，线段函数记做$L(x)$，则对$[a,b]$区间上任意一点x，有 $\left|f(x)-L(x)\right| \le \dfrac{\max\limits_{a\le x\le b}\left|f''(t)\right|}{2}\left|(x-a)(x-b)\right|$。

线性插值的误差的上界与目标函数的二阶导数绝对值最大值成正比，考虑

$$|(x-a)(x-b)| \le \frac{1}{4}(a-b)^2$$

因此

$$|f(x)-L(x)| \le \frac{1}{8}\mathrm{M}_2(a-b)^2$$

其中 M_2 表示在[a,b]区间内函数二阶导数绝对值的最大值。那么要想控制线性插值的误差，就需要根据 M_2 来调整插值点的间隔。设置间隔为

$$\mathrm{gap} = \frac{\mathrm{K}}{\sqrt{\mathrm{M}_2}}$$

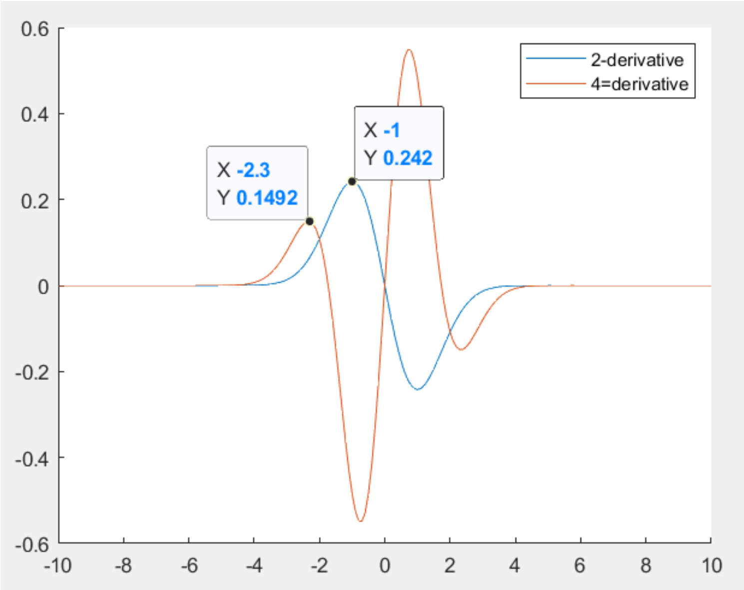
将间隔代入上述不等式，有

$$\mathrm{error}_{\mathrm{interp}} \le \frac{\mathrm{K}^2}{8}$$

令 $\mathrm{K}^2/8$=1e-6，解得$\mathrm{K} = \sqrt{8\times 10^{-6}}$

对于函数

$$\Phi''(x) = \frac{-1}{\sqrt{2\pi}}x\mathrm{e}^{-\frac{1}{2}x^2}$$



鉴于 M_2 极值点在 x=-1，因此在[-1.1,0]内，二阶导数绝对值最大值 M_2 可以取为全局最大值；在[-13,-1.1)内 M_2 可以取小区间[a,b]的右端点处的二阶导数绝对值。这种处理一方面能够在较长的单调区间内有效减少插值点的个数，另一方面实现起来较为简单，且不会损害算法的准确度。

选好 M_2 之后，可以利用下述公式计算间隔，得到插值点的向量。

$$\mathrm{gap} = \frac{\mathrm{K}}{\sqrt{\mathrm{M}_2}}$$

- **关于数值积分的误差控制**

根据数值积分误差公式，

$$|I - T_n| \leqslant \frac{h^2}{12} M_2 (b-a).$$

$$|I - S_n| \leqslant \frac{h^4}{180} M_4 (b-a),$$

要想控制误差上界为 1e-6，就有数值积分步长（t 代表梯形，s 代表 simpson）：

$$h_t = \sqrt{\frac{12 \times 10^{-6}}{13 M_2}}$$

$$h_s = \sqrt[4]{\frac{180 \times 10^{-6}}{13 M_4}}$$

M_2 的选取方法上面已有说明。

鉴于 M_4 的次极值点在 x=-2.3344，因此在[-2.4,0]内，四阶导数绝对值最大值 M_4 可以取为全局最大值；在[-13,-2.4)内 M_4 可以取小区间[a,b]的右端点处的四阶导数绝对值。

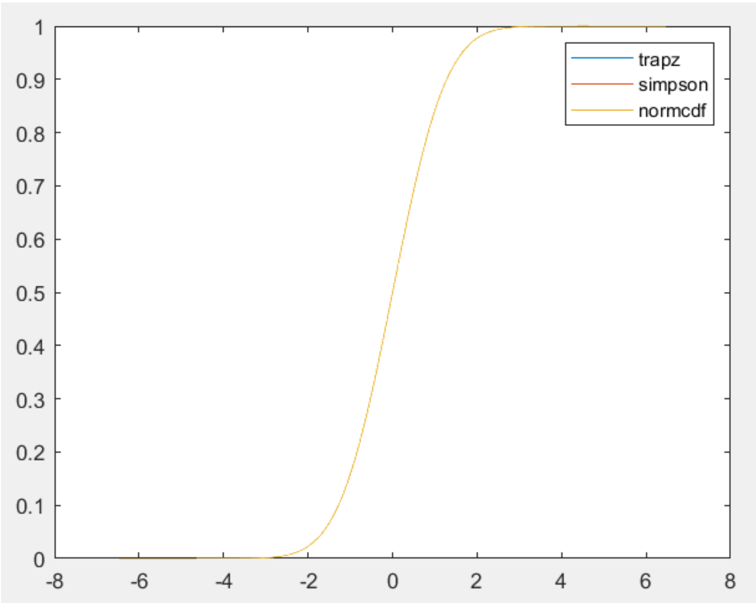
这种处理可以在 M_2、M_4 较小的区间内实现更少的数值积分等分数，同时实现起来较为简单，且不影响算法的精确度。

补充一点，如果不采用上述步长确定方法，而是直接给定一个非常大的等分数，会导致计算时间过长，不推荐采用。

- **数值积分**

根据上面所阐述的方法，对于插值点的选取、数值积分的等分数进行优化。最终插值点在[-13,13]上一共有 861 个，数值积分最大步长为 trapz：0.002、simpson：0.0708。

下图呈现了数值积分在插值点处的结果平滑连线，以及和 normcdf 结果的对比。直观上对比不出差异，说明误差很小。



取插值点向量 xinterp=-13:1e-5:13，采用 interp1(x 向量,函数值向量,插值点,'linear')的方法计算插值结果，并与 normcdf 结果进行对比。引入 quad、quadl 之后，下表展示了最大绝对误差的情况：

	积分误差	插值后误差
Trapz~1e-6	8.090626070507270e-08	1.076927567067454e-06
Simpson~1e-6	4.944393339201270e-09	9.960880872350586e-07
Quad~1e-6	0.001568018637123	0.001567846397959
Quadl~1e-6	1.074842574017715e-06	1.602251848292075e-06
Integral~1e-6	4.440892098500626e-16	9.999901545154621e-07

发现 quad 方法虽然指定了误差容限为 1e-6，但是计算精度仍然不能达到要求。换用 integral（全局自适应积分）之后误差大大减小。

综上，simpson 算法插值后误差最小，与全局自适应 integral 的结果相当，而梯形算法 trapz、自适应 Guass-Lobatto 法 quadl 次之。自适应 simpson 公式误差较大，不推荐采用。

- **参考资料**

1. MATLAB 函数参考页：

   trapz、quad、quadl、integral、interp1、fminbnd、fliplr（翻转向量）

2. 课程参考代码：

   simpson.m

3. 《大学数学实验_姜启源_第二版》
4. 数学实验课程讲义_第二讲_数值积分