数学实验第八次实验报告

计 76　张翔　2017011568

2020 年 5 月 12 日

1　实验目的

1. 掌握概率统计的基本概念及用 MATLAB 实现的方法;
2. 练习用这些方法解决实际问题。

2　Ch11-P6 炮弹问题

2.1　问题分析与模型建立

设目标中心为坐标原点，则圆形区域可以表示为

$$\Omega:\, x^2+y^2\le a^2$$

半径 $a=100$。根据题意，弹着点 $(x,\,y)$ 是符合二维正态分布的随机变量，概率密度函数为

$$f(x,\,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\exp\left(-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_x^2}-2r\frac{xy}{\sigma_x\sigma_y}+\frac{y^2}{\sigma_y^2}\right)\right)$$

其中 $\sigma_x=80,\,\sigma_y=50,\,r=0.4$。则炮弹命中圆形区域的概率为下列二重积分

$$P=\iint_\Omega f(x,\,y)\mathrm{d}x\mathrm{d}y$$

2.2　算法设计

上述积分难以直接计算，可以使用课上介绍的 Monte Carlo 方法进行估计。其中一个方法是均值估计法，对于 $[-a,\,a]$ 上均匀分布的随机变量 $X,\,Y$ 进行采样，可以得到积分的计算式为

$$P=\iint_\Omega f(x,\,y)\mathrm{d}x\mathrm{d}y\approx\frac{(2r)^2}{n}\sum_{k=1}^{m}f(x_k,\,y_k)$$

其中 $(x_k,\,y_k)$ 是总共 n 个随机变量 $(X,\,Y)$ 采样点中落在 Ω 区域内的 m 个点。生成均匀分布的随机数可以使用 MATLAB 的 `unifrnd` 函数实现。

另外，由于 Matlab 中自带生成二维正态分布向量的工具，可以使用它来进行随机投点，生成 n 维满足给定分布的向量，如果其中有 m 个元素在 Ω 内，则积分为

$$P=\frac{m}{n}$$

注意到被积函数关于原点中心对称，因此积分式可以写成下面的二次积分形式

$$P=2\iint_{\Omega\cap\{y>=0\}}f(x,\,y)\mathrm{d}x\mathrm{d}y=2\int_{-a}^{a}\int_0^{\sqrt{a^2-x^2}}f(x,\,y)\mathrm{d}x\ \mathrm{d}y$$

上述积分也可以使用 Matlab 的数值积分工具 `integral2` 计算。最后可以将随机模拟的结果与直接数值计算得到的结果进行比较。

2.3 Matlab 程序

随机投点法

```
%% data
a = 100;
sx = 80;
sy = 50;
r = 0.4;
n = 1000000;
%% sampling
sigma = [sx^2, sx*sy*r; sx*sy*r, sy^2];
x = mvnrnd([0, 0], sigma, n);
m = 0;
for i = 1:n
    if x(i,1)^2 + x(i,2)^2 <= a^2
        m = m + 1;
    end
end
P1 = m / n;
```

均值估计法

```
%% data
a = 100;
sx = 80;
sy = 50;
r = 0.4;
n = 100000;
%% sampling
x = unifrnd(-a, a, 1, n);
y = unifrnd(-a, a, 1, n);
m = 0; z = 0;
coeff = 1 / (2 * pi * sx * sy * sqrt(1 - r^2));
for i = 1:n
    if x(i)^2 + y(i)^2 <= a^2
        f = coeff * exp(-1/(2*(1-r^2)) * (x(i)^2/sx^2 - 2*r*x(i)*y(i)/sx/
            sy + y(i)^2/sy^2));
        z = z + f;
        m = m + 1;
    end
end
P = 4 * a^2 * z / n;
```

直接数值计算

```
a = 100;
sx = 80;
sy = 50;
r = 0.4;
coeff = 1 / (2 * pi * sx * sy * sqrt(1 - r^2));
fun = @(x, y) coeff * exp(-1/(2*(1-r^2)) * (x.^2/sx^2 - 2*r*x.*y/sx/sy + y.^2/sy^2));
ymax = @(x) sqrt(a^2 - x.^2);
P2 = 2 * integral2(fun, -a, a, 0, ymax);
```

2.4 计算结果与分析

使用直接数值计算的方法可以得到结果为 $P = 0.697939$。

通过修改模拟次数 n，多次求解后可以得到随机模拟方法的结果，其中每个 n 下随机模拟 10 次，取结果的平均值得到下列表格（下列误差是相对于直接数值计算的结果而言的）

表 1: 不同 n 值下的计算结果

	均值估计法			随机投点法		
n	均值	方差	误差	均值	方差	误差
10^4	0.69890	3.2746e-5	2.7631e-3	0.69971	3.7596e-6	3.4407e-3
10^6	0.69792	2.5733e-7	3.5206e-5	0.69790	4.1423e-7	1.3547e-4
10^8	0.69795	6.5695e-9	8.6054e-6	0.69793	2.0897e-9	1.0941e-5

由上表可以看出，随着模拟次数 n 的增加，结果的方差与误差均减小，采样精度则更高，且容易发现当 n 较大时，误差的减小速度变慢，如比较三组 n 下随机投点法得到的误差，可以发现它们的减小速度大约是 $n^{1/2}$ 阶，符合课本中关于 Monte Carlo 方法精度的结论以及大数定律。

另外，在相同的 n 下，均值估计法在 n 较小时没有随机投点法稳定 (方差更大)，但后两组较大的 n 下则比后者更稳定，且总体相对于精确值的误差比后者更小，因此，在本题的条件下，使用均值估计法是更优的。

2.5 结论

根据计算结果进行估计，可得炮弹命中圆形区域的概率为 0.698。

3 Ch11-P8 报童问题

3.1 问题分析与模型建立

由于每天报纸的需求量是随机的，报童每天利润也应是一个随机变量。记每天购进报纸 n 份，需求量 r 的概率为 $p(r)$，购进、零售、退回价格分别为 a, b, c，考虑未卖完和完全卖完两种情况，每天利润如下

$$V(n) = \sum_{r=0}^{n-1} [(b-a)r - (a-c)(n-r)]\, p(r) + \sum_{r=n}^{+\infty} [(b-a)n]\, p(r)$$

购进价格为购买量 n 的函数

$$a = a(n) = A(1 - \frac{n}{K})$$

报纸的日需求量可以通过统计历史数据得到，它是离散的，但根据生活经验，r 一般比较大，可以看作是连续随机变量 x，对于购入量 n 同理。根据题意，$x \sim N(\mu, \sigma^2)$。因此可以将日利润写作

$$V(n) = \int_0^n \left[(b-a(n))\,x - (a(n)-c)\,(n-x)\right] f(x)\mathrm{d}x + \int_n^{+\infty} \left[(b-a(n))\,n\right] f(x)\mathrm{d}x$$

其中 $f(x)$ 为 $N(\mu, \sigma^2)$ 的概率密度函数。

上式对 n 求导可得

$$V'(n) = (b-a(n))nf(n) + \int_0^n \left[c - a'(n)n - a(n)\right] f(x)\mathrm{d}x - (b-a(n))nf(n) + \int_n^{+\infty} \left[b - a'(n)n - a(n)\right] f(x)\mathrm{d}x$$

代入 $a(n), a'(n)$ 可得

$$V'(n) = \int_0^n \left[c + A\left(\frac{2n}{K} - 1\right)\right] f(x)\mathrm{d}x + \int_n^{+\infty} \left[b + A\left(\frac{2n}{K} - 1\right)\right] f(x)\mathrm{d}x$$

以及

$$V''(n) = (c-b)f(n) + \frac{2A}{K}\int_0^{+\infty} f(x)\mathrm{d}x$$

令 $V'(n) = 0$，考虑到题给的均值 μ 比标准差 σ 大得多，可以作如下的近似

$$\int_0^n f(x)\mathrm{d}x \approx \int_{-\infty}^n f(x)\mathrm{d}x$$

代入可得

$$\int_{-\infty}^n f(x)dx = \frac{b + A\left(\frac{2n}{K} - 1\right)}{b-c} \tag{1}$$

根据上式得到的 n 可能是问题的最优解，但还需要验证 $V''(n) < 0$，这一步可以在计算出 n 后进行处理。

3.2 算法设计

上述模型得到的 (1) 式是一个关于 n 的方程，它难以直接给出显式解，可以使用 MATLAB 的 `fzero` 对它进行数值求解。为了去掉积分符号，可以使用 `norminv` 得到正态分布的逆分布函数，以建立最终的方程进行求解。此时目标方程变为

$$F^{-1}\left(\frac{b + A\left(\frac{2n}{K} - 1\right)}{b-c}\right) - n = 0$$

解出 n 后代入 $V''(n)$，验证 $V''(n) < 0$ 时即说明此时最优。

3.3 Matlab 程序

计算得到 n 后，由于 $V''(n) < (c-b)f(n) + \frac{2A}{K}$，只需要验证后者小于 0 即可，无需计算正态分布的累计分布函数。

```
K = 50000;
[x,fval,exitflag,output] = fzero(@(x)eqn(K, x), [200 3000]);
val = validate(K, x);
if val < 0
    fprintf('Found best solution\n');
else
```

```
    fprintf('Not best solution\n');
end

%% plot various K
krange = 25000:1000:1000000;
res = [];
for K = krange
    [t,fval,exitflag,~] = fzero(@(x)eqn(K, x), [200 3000]);
    if exitflag ~= 1
        warning('Failed to converge\n');
    end
    res = [res t];
end
plot(1./krange, res), xlabel('1/k'),ylabel('Best N');

function y = eqn(K, n)
    mu = 2000;
    sigma = 50;
    A = 0.5;
    b = 0.5;
    c = 0.35;
    tmp = (b + A * (2 * n / K - 1)) / (b - c);
    y = norminv(tmp, mu, sigma) - n;
end

function y = validate(K, n)
    mu = 2000;
    sigma = 50;
    A = 0.5;
    b = 0.5;
    c = 0.35;
    y = (c - b) * normpdf(n, mu, sigma) + 2 * A / K;
end
```

3.4 计算结果与分析

在题给的 $K = 50000$ 条件下，求得 $n = 1.9682 \times 10^3$，取整数得 $n = 1968$，此时 $V''(n) < -9.58 \times 10^{-4} < 1$，因此该点是最优解。

注意到取最优值时 n 满足的 (1) 式与课本上的报童问题比较类似，它们的分母均为 $b - c$，而分子的 a 在此处对应 $A(1 - \frac{2n}{K})$，这个式子与实际购入价格的表达式相似，仅仅是在 $\frac{n}{K}$ 项前多了系数 2。课本上对于 a, b, c 参数的讨论比较详细，这里就不赘述了，而是主要讨论一下参数 K 对结果的影响。

直观来看，题目所给的购入价方程的成立需要满足 $n \leq K$，这个 K 可以理解为发行商对于客户的最大购买需求的一个预估值，发行商置价格的衰减系数 $1/K$ 与 K 成反比。考虑 $K \in [25000, 1000000]$，作出最佳购入量 n 与价格衰减系数 $1/K$ 的图线如下

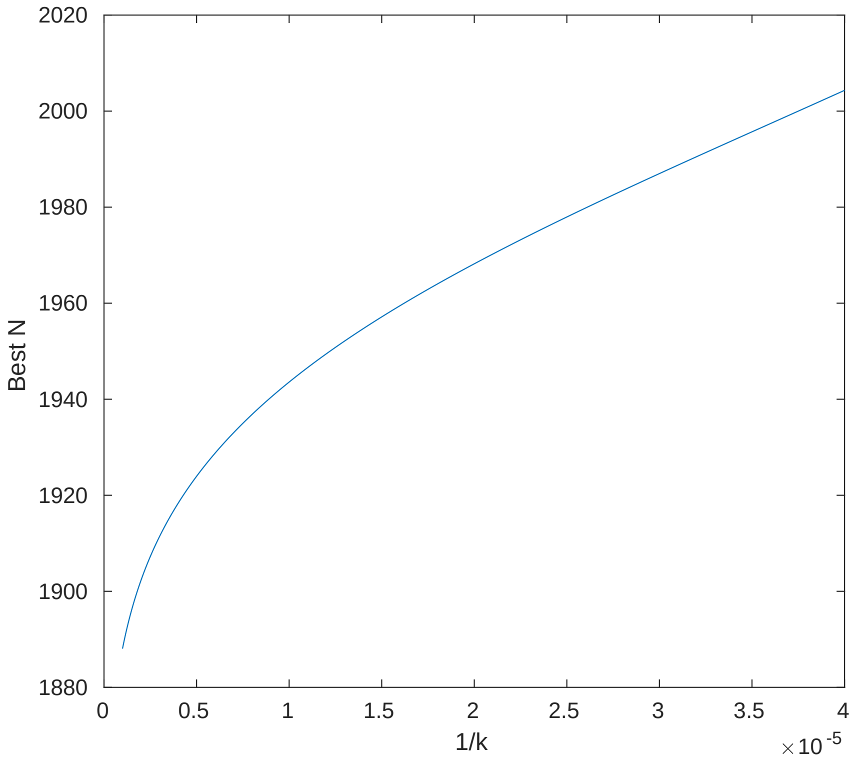

图 1: 最佳购入量 n 与价格衰减系数 $1/K$ 的关系

从图中可以看出，当衰减系数 $1/K > 2 \times 10^{-5}$ 时，即 $K < 50000$ 时，最佳购入量与 $1/K$ 有近似线性的关系，使用最小二乘拟合得到

$$\hat{n} = 1.797 \times 10^6 \times \frac{1}{K} + 1933$$

$R^2 = 0.9995$，说明二者确实具有较好的线性性质。

而 $1/K < 2 \times 10^{-5}$ 时，使用 MATLAB 的拟合工具箱可以得到

$$\hat{n} = 1787 \left(\frac{1}{K}\right)^{0.2218} + 1805$$

$R^2 = 0.9999$，拟合程度良好。

因此，如果发行商的 K 会随着市场动态调整，可以使用上面的拟合函数对最优购入量进行估计。同时，可以看出，价格衰减系数足够大（或 K 较小，购入优惠较大）时，最佳购入量与优惠幅度是线性关系；而优惠较小时，最佳购入量将加速递减，在图像中的体现是 $1/K$ 在 0 附近时图线变得更加陡峭。题目所给的 K 值恰好是这两种情况的分界。

另外，题目中所给的购入价格定义式仅对一般情况有效，如果某些报童希望购入的数量超出了出版商的预期购买量 K，可能会出现购入价格为负数的情况，这是不符合日常规律的。可以考虑为定义式增加约束，如购入量超出一定值后价格不变等，使模型更加合理。

3.5 结论

$K = 50000$ 时，报童每天买入 1968 份报纸，可获得的期望利润最大。

4 Ch11-P10 轧钢问题

4.1 问题分析与模型建立

根据题意，如果轧钢设备设定的平均长度过大，则精轧时需要切割的概率更大，造成很多浪费；反之，如果平均长度过小，虽然额外切割带来的浪费减少，但后续报废的可能性增加。最佳的选择需要权衡粗轧平均长度 m，使得总浪费最少。

设粗轧后的钢材长度为 x，这个随机变量 $X \sim N(m, \sigma)$，其中 m 可控，σ 取决于设备，无法改变。X 的概率密度函数记作 $f(x)$，那么每粗轧一根钢材期望浪费为 W_1，其中有两部分，分别为精轧时需要切割的量、整根报废的浪费量，如下：

$$W_1 = \int_l^{+\infty} (x-l) f(x) \mathrm{d}x + \int_{-\infty}^{l} x f(x) \mathrm{d}x$$

每得到一根规定长度钢材的期望浪费是一个条件期望问题，即

$$E(\{\text{粗轧一根的浪费量}\} \mid \{\text{得到一根规定长度钢材}\}) = \frac{W_1}{P\{\text{得到一根规定长度钢材}\}}$$

记这个条件期望为 W_2，则

$$W_2 = \frac{W_1}{\int_l^{+\infty} f(x) \mathrm{d}x}$$

记不报废的概率为 $P = \int_l^{+\infty} f(x) \mathrm{d}x$，上述表达式可以化为

$$W_1 = m - lP, \ W_2 = \frac{m}{P} - l$$

4.2 算法设计

虽然上述 $W_1.W_2$ 的表达式较为简洁，但由于 $P(m) = \int_l^{+\infty} f(x) \mathrm{d}x$ 表达式较为复杂，直接求导得极值的方法较为繁琐，可以考虑使用数值方法计算最小值点（如 `fminbnd`）。

4.3 Matlab 程序

```matlab
%% input data
sigma = 0.2;
l = 2;

%% plot
xrange=[2 3];
hold on,
fplot(@(x)w1(sigma, l, x), xrange),
fplot(@(x)w2(sigma, l, x), xrange),
legend('W_1', 'W_2');

%% solve
[x1,fval1,exitflag1,output1] = fminbnd(@(x)w1(sigma, l, x), xrange(1),
    xrange(2));
[x2,fval2,exitflag2,output2] = fminbnd(@(x)w2(sigma, l, x), xrange(1),
    xrange(2));

```

```
%% objectives
function y = w1(sigma, l, m)
    P = (1 - normcdf(l, m, sigma));
    y = m - l * P;
end

function y = w2(sigma, l, m)
    P = (1 - normcdf(l, m, sigma));
    y = m ./ P - l;
end
```

4.4 计算结果与分析

首先作出 $W_1(m), W_2(m)$ 的图像，以确定最小值所在区间

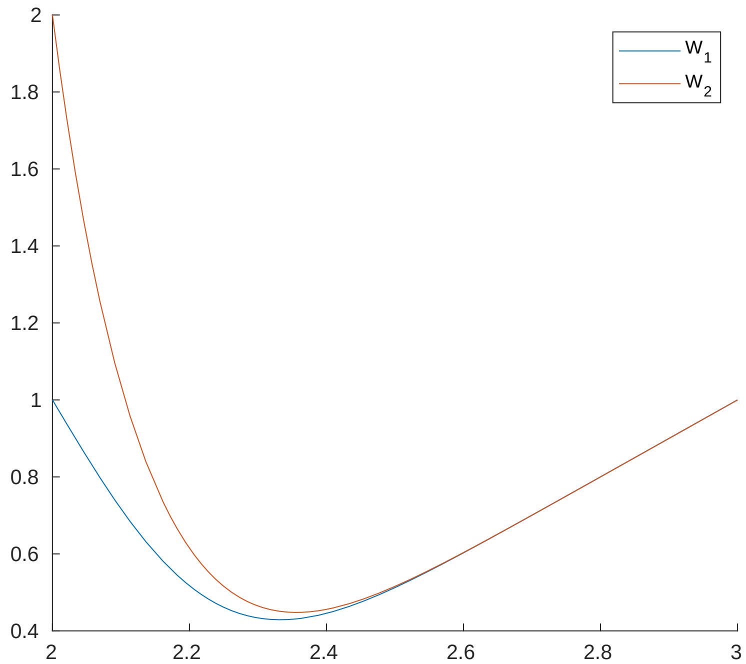

图 2: $W_1(m), W_2(m)$ 图像

可以看出，最小值均在 [2, 3] 内，因此使用 `fminbnd` 时传入该区间作为参数。

计算得 W_1 取最小值时，$m = 2.3327$，此时 $W_1 = 0.4289$；W_2 取最小值时，$m = 2.3562$，此时 $W_2 = 0.4479$。

由上述结果容易得到，当 m 减小或增大时，两种情况的浪费均增加，且当 m 减小时的浪费增长速度比 m 增大时的更快，且 m 减小时 W_1 与 W_2 的差距显著增大。这说明在精轧期间，整根报废带来的浪费要比切割多余的要更多一些。而 m 减小时，钢材达到规定长度的概率减小，合格率更低，这种代价分摊到合格钢管上将使得期望浪费更高；m 增大时，由于合格率显著提升，W_1 与 W_2 的差距已经不明显。

如果将上述 m 代入累积分布函数计算，可以得到两种情况下，粗轧的钢管长度超过 2m（不报废）的概率分别为 95.19% 与 96.25%，显然这是偏向于减少整根报废的策略。

4.5 结论

当目标为每粗轧一根钢材的浪费最小时，设定均值 $m = 2.33$ m；当目标为每得到一根规定长度钢材的浪费最小时，设定均值 $m = 2.36$ m。

5 收获与建议

通过这次实验，我对使用 MATLAB 求解统计问题的方法有了更深刻的理解，并复习了一些概率论与数理统计中的重要知识。希望老师能够在课堂上提供更多常用的与统计问题相关的例子。