西安电子科技大学

XIDIAN UNIVERSITY

概率论

概率论课程实践报告

陈德创

19030500217

计算机科学与技术专业

指导教师 史罡

December 15, 2020

概率论课程实践报告

陈德创　　　　19030500217

西安电子科技大学

日期：2020 年 12 月 15 日

目录

1 实验内容

1). 利用均匀分布 $[-1,1]$，使用中心极限定理产生正态分布的随机数 1000 个，其中每次用于产生正态分布随机数的均匀分布样本容量 $n \geq 50$

2). 在产生的 1000 个随机数中，均匀随机的抽取一个容量为 100 的样本，选择有效的点估计量给出基于该样本的总体均值和方差的点估计值，并将其与理论值进行比较

3). 基于抽取的容量为 100 的样本给出总体均值和方差的置信度为 0.95 的置信区间，并考察总体参数是否落在置信区间中

2 正态分布的产生与样本的抽取

中心极限定理实际上是对相互独立且同分布的随机变量（设为 $X_1, X_2, ..., X_n$）的随机变量函数

$$Y_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \tag{1}$$

的分布情况的刻画。

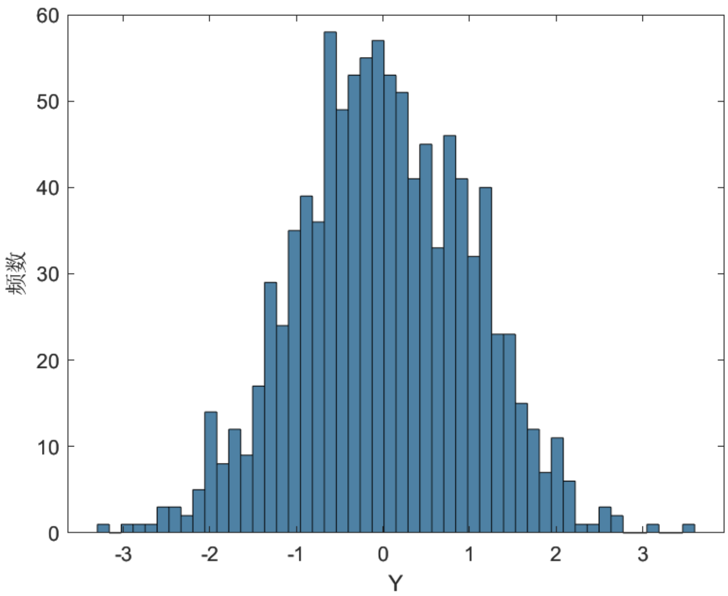
利用 *matlab* 中内置了 *rand* 函数，可以产生给定维数矩阵的 $[0,1]$ 区间的均匀分布。实际上我们并不关系这些变量具体的值，而只关心他们生成的高斯变量 Y。所以我们构建函数 $getGaussianVariable(n)$，接收均匀分布的样本容量 n，随机生成样本，利用公式 (1) 得到高斯变量并返回。

该函数的具体实现如下：

```
% n: 样本空间大小
function res = getGaussianVariable(n)
  t = rand(1, n) * 2 - 1;
  res = sum(t) / sqrt(n / 3);
end
```

利用此函数，我们可以得到一个满足标准正态分布且样本容量为 1000 的总体样本 $Y_1, Y_2, ..., Y_{1000}$，下图展示了利用 *histogram* 函数描绘的生成 1000 了个体的分布情况，可以看到样本大致符合正态分布。

图 1: 随机变量 Y 的频数分布直方图



我们需要在其中随机抽取 100 个个体组成样本。然而，*matlab* 中并没有给定我们一个直接的函数可

以在总体样本空间中抽取一定个体的函数，于是我们转变思路。考虑 $randperm(n)$ 函数，此函数将会等概率随机生成一个长度为 n 的排列，我们截取这个排列的前 100 项，将位于这些位置的个体抽取出来组成样本。

函数 $drawSample$ 的具体实现如下：

```
% x: 总体样本, n: 随机抽取的个体个数
function [data] = drawSample(x, n)
  len = length(x);
  randrk = randperm(len);
  data = zeros(1, n);
  for i = 1 : n
    data(i) = x(randrk(i));
  end
end
```

3 总体均值和方差的理论值以及点估计值的计算

由于总体遵循正态分布，只有 μ, σ^2 为未知参数，我们通过矩估计法可以方便快速地求出总体均值和方差地点估计值。计算公式如下：

$$\begin{aligned}\hat{\mu} &= \bar{X},\\ \hat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\end{aligned} \tag{2}$$

总体的理论均值和方差可由公式：

$$\begin{aligned}\mu &= \bar{Y},\\ \sigma^2 &= EY^2 - (EY)^2\end{aligned} \tag{3}$$

求得。

计算总体均值和方差的理论值以及基于所抽取的样本的总体均值和方差的点估计值代码实现如下：

```
% x为总体, data为样本
% 计算均值和方差的理论值
Ex = sum(x) / len;
Dx = sum(x.*x) /len - Ex^2;
display([Ex, Dx]);
% 计算均值和方差的点估计值
Edata = sum(data) / n;%均值
Ddata = sum((data - Edata).^2) / n;%方差
display([Edata, Ddata])
```

得到结果理论值 $Ex = 0.0091$, $Dx = 0.9988$，点估计值 $Edata = 0.1278$, $Ddata = 1.1267$。

4 总体均值和方差置信区间的计算

要求置信度为 0.95 的置信区间，即 $\alpha = 0.05$。

4.1 总体均值置信区间的计算

由课本定理可知，当总体 $X \sim N(\mu,\sigma^2)$，$X_1, X_2, ..., X_n$ 为来自总体 X 的一个样本，样本均值为 $\bar{X}$，则均值 μ 的置信水平为 $1-\alpha$ 的置信区间为

$$(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}) \tag{4}$$

其中，$\bar{X}$ 我们先前已经算出，那么只需要计算 $\frac{\sigma}{\sqrt{n}} z_{\frac{\alpha}{2}}$ 即可。其中唯一的未知量为 $z_{\frac{\alpha}{2}}$，查表可得 $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$，那么均值置信区间的计算就变得十分简单了，代码如下：

```
% 计算置信区间
z025 = 1.96;
% E的置信区间
deltaE = sqrt(Ddata) / sqrt(n) * z025;
disp("E的置信区间");
display([Edata - deltaE, Edata + deltaE]);
```

在 *matlab* 中运行上述代码我们可以得到均值 μ 的置信区间为 $[-0.0803, 0.3358]$，而总体均值的理论值为 0.0091，在置信区间内。

4.2 总体方差置信区间的计算

同样地，方差 σ^2 置信区间的计算，由课本定理可知当总体 $X \sim N(\mu,\sigma^2)$，其中 μ 已知，$X_1, x_2, ..., X_n$ 为来自总体 X 的一个样本，则方差 σ^2 的置信水平为 $1-\alpha$ 的置信区间为

$$\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)}, \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)} \right) \tag{5}$$

我们需要得到 $\chi^2_\alpha(n)$ 的值。由课本可知，费歇曾证明过，当 n 较大时，我们有

$$\chi^2_\alpha(n) \approx \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2 \tag{6}$$

通过式 (6)，我们可以得到 $\chi^2_{0.025}(100)$ 和 $\chi^2_{0.975}(100)$ 的近似值分别为 $129.0700, 73.7716$，再通过式 (5) 即可得到 σ^2 的置信区间。代码如下：

```
% 计算置信区间
z025 = 1.96;
kf025 = 0.5 * (z025 + sqrt(2 * n - 1)) ^ 2
kf975 = 0.5 * (-z025 + sqrt(2 * n - 1)) ^ 2
% D的置信区间
t = data - Edata;
LD = sum(t .* t) / kf025;
RD = sum(t .* t) / kf975;
disp("D的置信区间");
display([LD, RD]);
```

通过 *matlab* 运行上述代码可得到方差 σ^2 的置信区间为 $[0.8730, 1.5273]$，而总体方差的理论值为 1.0144，在区间内。

5 个人总结

很开心自己认真完成了这次概率论大作业，令我受益良多。

“纸上得来终觉浅，绝知此事要躬行”，文字的局限难免令我们难以理解知识的本质特征，而沦为空洞刻板，不知所以然的公式定理。通过这次作业，我对课本五到七章内容有了一个全面、直观的理解，这绝对不是单纯看课本或者刷题可以如此迅速达到的，可能这就是所谓“实践出真知”吧。

附录

A matlab 源码

```
% getGaussianVariable.m文件
% n: 样本空间大小
function res = getGaussianVariable(n)
  t = rand(1, n) * 2 - 1;
  res = sum(t) / sqrt(n / 3);
end

% drawSample.m文件
% x: 总体样本, n: 随机抽取的个体个数
function [data] = drawSample(x, n)
    len = length(x);
    randrk = randperm(len);
    data = zeros(1, n);
    for i = 1 : n
         data(i) = x(randrk(i));
    end
end

% Probability_theory.mxl文件
len = 1000;
% x用于储存利用中心极限定理生成的高斯分布
x = zeros(1, len);
% 获得一千个高斯变量
for i = 1 : len
    %x(i) = rand(1, 1)
    % 在[-1,1]的均匀分布取100个点, 得到高斯变量
    x(i) = getGaussianVariable(100);
end
histogram(x, 50);
ylabel('频数');
xlabel('Y')

n = 100;
% 在x中随机取n个点作为样本
data = drawSample(x, n);
display(data);
% 计算均值和方差的理论值
Ex = sum(x) / len;
Dx = sum(x.*x) /len - Ex^2;
disp("总体均值和方差的理论值");
display([Ex, Dx]);
% 计算均值和方差的点估计值
Edata = sum(data) / n;%均值
Ddata = sum((data - Edata).^2) / n;%方差
disp("总体均值和方差的点估计值");
display([Edata, Ddata])

% 计算置信区间
z025 = 1.96;
```

```
kf025 = 0.5 * (z025 + sqrt(2 * n - 1)) ^ 2
kf975 = 0.5 * (-z025 + sqrt(2 * n - 1)) ^ 2
% E的置信区间
deltaE = sqrt(Ddata) / sqrt(n) * z025;
disp("E的置信区间");
display([Edata - deltaE, Edata + deltaE]);
% D的置信区间
t = data - Edata;
LD = sum(t .* t) / kf025;
RD = sum(t .* t) / kf975;
disp("D的置信区间");
display([LD, RD]);
```

B matlab 运行界面

图 2: matlab 运行界面截屏