期末考试题型说明

一、选择题（20空，40分）

包含各章节的基本概念，NS-3中的基本概念（node，device，channel等）、使用方法、配套工具（如waf、gnuplot、netanim、brite等）、实验二涉及的模型（three-gpp-http, propagation-loss, random-walk等）。

二、判断题（10题，10分）

考查内容同上。

三、填空题（15空，15分）

考查重点概念：如离散事件仿真策略中的三阶段法、蒙特卡洛仿真的原理和步骤、随机变量产生的一般性方法及其适用条件、从数据到模型的理论建模方法、业务源模型建模的共同特征、移动性模型及其分类、无线链路信噪比计算等。

四、程序填空题（15空，15分）

示例：

1.下面产生瑞利分布随机数，其分布函数为

$$F(x)=1-e^{-\frac{x^2}{2\sigma^2}}\quad x>0$$

#include \<cstdlib>

#include \<iostream>

....

double Rayleigh(double sigma, double state[ ]) //产生一个瑞利分布随机数

{

double u1;

u1=__(3)__; //生成均匀分布随机数

return__(4)__; //生成瑞利分布随机数

}

....

2.下面为一段NS3程序，用于模拟一对节点之间的传输，其中一端模拟一个回声服务器，另一端模拟一个客户端，请把程序中的空补充完整。

#include "ns3/__(9)__.h" //内核模块

#include "ns3/network-module.h"

....

NodeContainer__(10)__;

nodes.Create (2);

PointToPointHelper pointToPoint;

pointToPoint.SetDeviceAttribute ("DataRate", StringValue ("5Mbps"));

pointToPoint.SetChannelAttribute ("Delay", StringValue ("2ms"));

NetDeviceContainer devices;

<u>（11）</u>= pointToPoint.Install (<u>（12）</u>); //在节点上安装网络设备

....

五、解答与设计题（2题，20分）

示例：
蒙特卡洛仿真中，稀有事件出现的概率、仿真量和仿真的精度（绝对精度和相对精度）之间有密切关系，请利用它们之间的关系解决以下问题：
(1)一个通信链路，设传输错误概率很小，如果在仿真中每观察到100个误码就进行一次误码率的统计，问得到的结果在95%置信度条件下的相对精度是多少？（5分）

α	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
$erf^{-1}(\alpha)$	0.0089	0.0177	0.0266	0.0355	0.0443	0.0532	0.0621	0.0710	0.0799	0.0889
$erfc^{-1}(\alpha)$	1.8214	1.6450	1.5345	1.4522	1.3859	1.3299	1.2812	1.2379	1.1988	1.1631

绪论

- 了解系统仿真的一般方法。
- 掌握蒙特卡洛仿真方法；
- 掌握网络仿真中关键模块的建模与仿真方法。
- 初步掌握NS-3等仿真软件的使用方法
- 平时成绩：20%
  实验成绩：40%（1个实验交报告，1个实验验收+报告）
  期末考试：40%

1-1 仿真概述

- 确定系统与随机系统
  - 确定系统

    > 活动完全确定性的系统。
    > 系统的输入一旦给定，系统内各个参量和系统输出随时间的变化关系都将完全确定。

  - 随机系统

    > 具有**随机性活动**的系统，随机型活动通常用**概率分布**加以描述。

- 模型

  > 为了便于对系统进行研究，对实际系统本质进行**抽象**、**简化**和**近似**。
  > 基于模型的实验：具有**时间短**、**成本低**的优点。

- 仿真

  > 组成：包括"**建模** - **实验** - **分析**"三个步骤
  > 计算机仿真：现代仿真技术中，把**数学模型**转化为**仿真模型**在计算机上进行的实验活动。

- 仿真的分类 按系统模型的特性分类
  - 模型的动态性
    - 静态仿真

> 模型本身**不随时间变化而变化**
> **不需要考虑时间因素**，即仿真模型中不考虑仿真时钟的管理与更新问题

- 动态仿真

> 模型本身的**各种行为都与时间密切相关**，针对这种模型的仿真都属于动态仿真。

- 状态变量变化的连续性
  - 连续系统仿真

> 系统本身的**状态变量是连续变化**的
> 往往可以建模为**微分方程**形式
> 在实际计算机处理时，需转换为**差分方程**的求解问题

  - 离散事件系统仿真

> 状态变量只在某些**离散的时间点**上发生变化
> 往往需要在统一的**仿真时钟控制**下仿真不同事件的处理过程，跟踪系统状态的变化，得到相应的输出。

- 模型的确定性
  - 确定性仿真

> 所研究的系统模型中所有元素的数学和逻辑关系都是**确定的**，不包含任何随机成分。

  - 随机性仿真

> 所研究的系统模型中包含**随机元素。**

- 系统模型的驱动方式
  - 数据流驱动

> 数据流从前端的实体流向后端的实体，**各个实体依次处理**，不需要考虑时间变量
> 适合于实体间**相互作用较为简单**的系统
> **单个通信链路**的仿真往往可以采用这种方式

  - 时间驱动

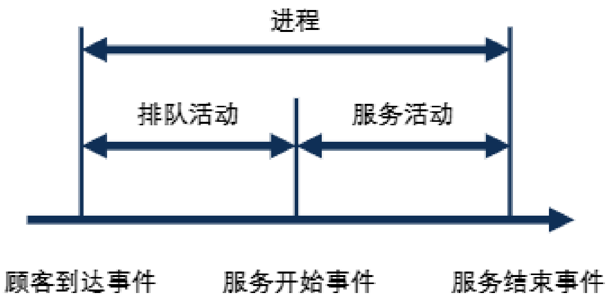
> 时间变量按一定步长推进， 整个仿真模型在**仿真时间变量**的驱动下工作。

  - 事件驱动

> 系统模型是离散事件系统模型，**仿真钟在事件控制下推进**

1-2 离散时间系统仿真

- 基本概念
  - 实体：在仿真中需要跟踪其行为的系统元素
  - 事件：引起系统状态发生变化的行为
  - 活动：表示两个**可以区分的事件之间的过程**，它标志着**系统状态的转移**
  - 进程：由若干个有序事件及若干个有序活动组成，一个进程描述了它所包括的事件及活动间的相互逻辑关系及时序关系。



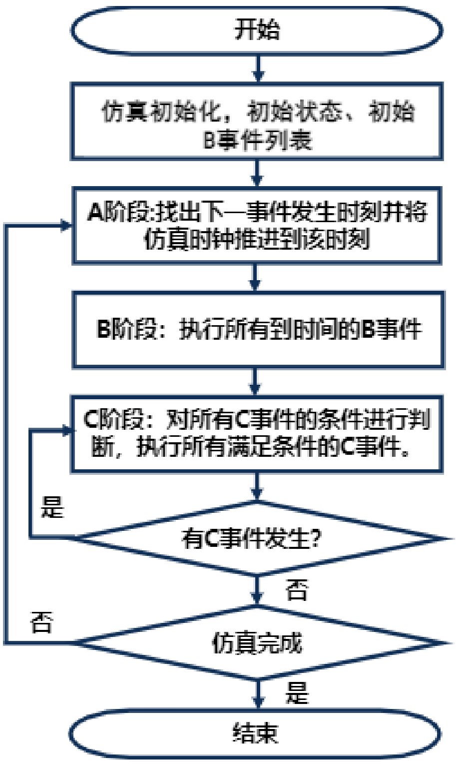
- 仿真时钟:

  > **表示仿真时间的变化**
  > 是仿真的时间基础、仿真过程的推进器和驱动器。

- 统计计数器:

  > 离散事件系统仿真的结果只有在**统计意义**下才有参考价值
  > 离散事件系统仿真模型中，需要有一个**统计计数**部件，以便统计系统中的有关变量。

- 三阶段法
  - 基本思想是**将无条件事件和有条件事件的处理相分离**，每一次修改仿真钟之后都会对两类事件分别进行处理。
  - B事件：发生时刻是**可预测**的，是**无条件**的
  - C事件：发生与否是**有条件**的，通常与时间没有直接关系
  - A阶段： **时间扫描。**
    推进仿真时钟。找出下一个最早发生的事件，将系统仿真钟推进到该事件的发生时刻。
    形成B事件列表。记录在该时刻要发生的B事件，形成一个B事件列表
  - B阶段:
    执行B事件列表
  - C阶段:
    查询C事件表
    逐一对事件表中的事件进行**条件测试**
    若条件满足则**执行相应的动作。**



蒙特卡洛方法

- 大数定理

  ➢ 蒙特卡洛方法可以看作由随机变量X的样本X_1，X_2，…，X_N的**算术平均值**:

$$\overline{X}_N=\frac{1}{N}\sum_{i=1}^{N}X_i$$

  作为所求解期望的近似值。由**大数定理**可知，如果X_1，X_2，…，X_N独立同分布，且具有有限期望值（$E(X)<\infty$），则

$$P\left(\lim_{N\to\infty}\overline{X}_N=E(X)\right)=1$$

  ➢ **即随机变量X的样本的算术平均值$\overline{X}_N$，当样本数N充分大时，以概率1收敛于它的期望值$E(X)$。**

- 中心极限定理

  ➢ 蒙特卡洛方法的**近似值与真值的误差**问题，概率论的中心极限定理给出了答案。该定理指出，如果随机变量序列X_1，X_2，…，X_N独立同分布，且具有有限非零的方差σ^2，即

$$0\neq\sigma^2=\int(x-E(X))^2f(x)dx<\infty$$

  当N足够大时

$$\frac{\sum_{i=1}^{N}X_i-NE(x)}{\sigma\sqrt{N}} \text{ 满足正态分布：} \mathcal{N}(0,1)$$

  整理得：

$$\frac{\sum_{i=1}^{N}X_i-NE(x)}{\sigma\sqrt{N}}=\frac{\sqrt{N}\ \left(\overline{X_N}-E(x)\right)}{\sigma}$$

  求解概率：

$$P(|\overline{X_N}-E(x)|<\frac{\lambda_\alpha\sigma}{\sqrt{N}})=P(|\frac{\sqrt{N}(\overline{X_N}-E(x))}{\sigma}|<\lambda_\alpha)$$

9

  ➢ 则当N充分大时，有如下的近似式：

$$P\left(\left|\overline{X}_N-E(X)\right|<\frac{\lambda_\alpha\sigma}{\sqrt{N}}\right)\approx\frac{2}{\sqrt{2\pi}}\int_0^{\lambda_\alpha}e^{-t^2/2}dt=1-\alpha$$

  - 其中1-α称为置信度。

$$\alpha=2Q(\lambda_\alpha)=\mathrm{erfc}(\frac{\lambda_\alpha}{\sqrt{2}})$$

  - 这表明，不等式

$$\left|\overline{X}_N-E(X)\right|<\frac{\lambda_\alpha\sigma}{\sqrt{N}}$$

  近似地以概率1−α成立，且误差收敛速度的阶为：$O(N^{-1/2})$

- 蒙特卡洛误差

$$\text{MC方法的误差：}\varepsilon=\frac{\lambda_\alpha\sigma}{\sqrt{N}}$$

  其中，λ_α 与 α 一一对应，查表就可得出。
  均方差 σ 可以计算出估计值

$$\sigma=\sqrt{\frac{1}{N}\sum_{i=1}^{N}X_i^2-(\frac{1}{N}\sum_{i=1}^{N}X_i)^2}$$

- 优点
  - 能够比较逼真地描述具有随机性质的事物的特点
  - 适合多维复杂问题
  - **收敛速度与问题的维数无关**
  - 具有同时计算多个方案与多个未知量的能力
  - **误差容易确定**
- 缺点
  - 收敛速度慢
  - 误差有随机性
- 蒙塔卡洛方法的一般步骤
  - 构造或描述概率过程
  - 从已知概率分布抽样

    > 所谓的从已知概率分布抽样指的是**随机试验过程**
    > 随机模型中必须要包含**某些已知概率分布的随机变量或随机过程**作为输入
    > 进行随机试验的过程就是对这些随机变量的样本或随机过程的**样本函数作为输入产生相应输出的过程**
    > 如何**产生已知分布的随机变量或随机过程**是蒙特卡洛方法中的一个关键问题

  - 建立估计量

    > 蒙特卡洛方法所得到的问题的解总是**对真实解的一个估计**

- 精度分析
  - 设数据的精确值是 x_0，通过仿真得到的估计值为 $\hat{x}$，$\hat{x}$ 是一个服从某种分布的随机变量。如果估计值 $\hat{x}$ 有概率 $1-\alpha$ 落在某一区间 $[x_0-\Delta, x_0+\Delta]$ 上，称区间 $[x_0-\Delta, x_0+\Delta]$ 为置信度 $1-\alpha$ 的置信区间
  - 绝对精度：Δ
  - 相对精度 $\frac{\Delta}{x0}$
  - 近似正态分布确定仿真次数：

    ➢根据大数定理，当试验次数 $n\to\infty$，试验中事件发生次数k服从均值为np方差为$np(1\text{-}p)$的正态分布，即

    $$p\left(\left|\frac{k}{n}-p\right|<\delta\right)\approx\frac{1}{\sqrt{2\pi}}\int_{-a}^{a}\exp\left(-\frac{x^2}{2}\right)dx=2\Phi(a)$$

    ➢其中

    $$a=\frac{n\delta}{\sqrt{np(1-p)}}$$

    $$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x\exp\left(-\frac{x^2}{2}\right)dx=\frac{1}{2}erf\left(\frac{x}{\sqrt{2}}\right)$$

➢这样，给定置信度$1-\alpha$和绝对精度以及事件的概率值p，就可以求解方程

$$erf\left(\frac{n\delta}{\sqrt{2np(1-p)}}\right)=1-\alpha$$

➢得出最小仿真次数n。如果概率值p未知，可用频率估计代替。

$\hat{x}=\frac{k}{n}$ 有概率 $1-\alpha$ 落在 $[x_0-\delta, x_0+\delta]$ 上.

由大数定律得 $k\sim N(np, np(1-p))$

所以 $\frac{k}{n}\sim N\left(p, \frac{np(1-p)}{n^2}\right)$

以均值 p 代替精确值 x_0

$$P\left(p-\Delta<\frac{k}{n}<p+\Delta\right)=\phi\left(\frac{p+\Delta-\mu}{\sigma}\right)-\phi\left(\frac{p-\Delta-\mu}{\sigma}\right)$$

$$=2\phi\left(\frac{n\Delta}{\sqrt{np(1-p)}}\right)-1$$

$$\begin{cases} erf(x)=1-2Q(\sqrt{2}x) \\ Q(x)=1-\phi(x) \end{cases} \Rightarrow \phi(x)=\frac{1}{2}+\frac{1}{2}erf\left(\frac{x}{\sqrt{2}}\right)$$

所以 $1-\alpha=erf\left(\frac{x}{\sqrt{2}}\right)$

即 $erf\left(\frac{n\Delta}{\sqrt{2np(1-p)}}\right)=1-\alpha$

- 近似泊松分布确定仿真次数：

$$1-\alpha=F(np+n\delta)-F(np-n\delta)$$

$F(x)$ 是参数为 np 的泊松分布的分布函数

- 相对精度确定仿真次数

➢在很多情况下，相对精度比绝对精度更有意义，因此在实际仿真中往往需要通过置信度和相对精度确定最小仿真次数。

➢给定仿真的相对精度要求 $r=\delta/p$，则 $\delta=pr$，将之代入到

$$erf\left(\frac{n\delta}{\sqrt{2np(1-p)}}\right)=1-\alpha$$

➢并整理得到已知相对精度条件下的最小仿真次数为

$$n=\frac{2(1-p)}{pr^2}(erf^{-1}(1-\alpha))^2$$

- 结论

> 对于稀有事件仿真统计而言，只有稀有事件出现次数大于100，才能将相对误差不超过20%的可能性控制在5%以内

> 在实际稀有事件仿真中，通常要求出错个数超过100

- 随机数生成算法
  |名称|生成器|多流和子流支持|全精度近似周期| |:--------:|:--------:|:--------:|:--------:| |mt19937ar| 梅森旋转| 否| $2^{19937}-1$| |mrg32k3a| 组合多递归生成器|是 | 2^{191} （2^{63} 个流，长度为 2^{127}） | |mcg16807| 乘法同余生成器| 否|$2^{31}-2$|
- 反变换法
  求分布函数的反函数作为生成函数
- 组合法

  组合法基本流程:

  (1) 产生随机整数I，使$P\{I=i\}=p_i$

  (2) 产生具有分布函数$F_i(x)$的随机变量 x_i

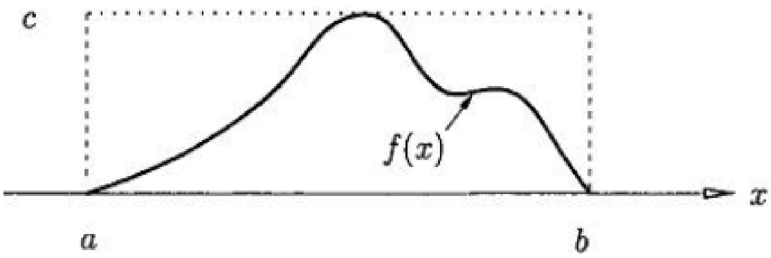
  (3) 令$x=x_i$
- 舍选法
  使用场景：目标连续随机变量的**累积分布函数**难以写成闭式表达式，但是其**概率密度函数**可以写成闭式表达式形式。

  ➢舍选法基本流程：　　a

  □产生(a,b)区间内均匀分布的随机变量u_1;

  □产生$(0,c)$区间内均匀分布的随机变量u_2;

  □如果 $u_2 \le f_X(u_1)$ 则 $X=u_1$；否则，回到第(1)步。

- 中心极限定理法产生高斯分布随机变量

  ➢中心极限定理指出，当 $N\to\infty$ 时，N个独立随机变量之和的分布趋近于高斯分布

  ➢基本步骤：

  □产生若干个（0,1）区间上的均匀分布随机数U_i

  □$U_i-\frac{1}{2}$服从（$-\frac{1}{2}$，$\frac{1}{2}$）的均匀分布，方差为$\frac{1}{12}$

  □N个$U_i-\frac{1}{2}$相加，近似服从均值为均值为0，方差为$\frac{N}{12}$的高斯分布

  □取$B=\sqrt{\frac{12}{N}}$，则$X=B\sum_{i=1}^{N}\left(U_i-\frac{1}{2}\right)$近似服从标准高斯分布

  □一般，取：$N=12$

  - 中心极限定理法的截尾现象

    > 按上述参数得到的随机变量只能在区间(-6,6)区间上取值
    > 利用中心极限定理方法产生高斯分布随机变量时，实际产生变量的范围总是无法达到

> **高斯分布的尾部**，这种现象称为截尾
> 增大N值，可以减小这种影响，却会增大计算量

  - 如果采用中心极限定理方法产生高斯分布，应该根据需要合理选取N值，以在**仿真速度和精度**之间作出折中

- Box-Muller方法产生高斯分布随机变量

➢根据瑞利分布和高斯分布之间的对应关系

□如果X和Y为均值为0、方差为σ^2的独立高斯随机变量，那么$R=\sqrt{X^2+Y^2}$服从参数为σ 的瑞利分布

□定义：把X和Y分别作为复平面的两个坐标，则相角θ服从$[0,2\pi]$上的均匀分布

➢瑞利分布的概率密度函数和累积分布函数

$$f_R(r)=\frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)u(r)$$

$$F_R(r)=\int_0^r\frac{y}{\sigma^2}\exp\left(-\frac{y^2}{2\sigma^2}\right)dy=1-\exp\left(-\frac{r^2}{2\sigma^2}\right),r\geq 0$$

➢瑞利分布的累积分布函数具有闭式表达式，可以采用反变换法产生瑞利分布随机变量

➢Box-Muller方法的步骤

□产生两个独立的(0,1)区间上均匀分布随机变量u_1和u_2；

□根据瑞利分布的累积分布函数，对u_1进行反变换

$$F_R(r)=1-\exp\left(-\frac{r^2}{2\sigma^2}\right),r\geq 0$$

□产生两个独立的、均值为0方差为1的高斯分布随机变量。

$$X=\sqrt{-2\ln u_1}\cos(2\pi u_2)$$

$$Y=\sqrt{-2\ln u_1}\sin(2\pi u_2)$$

- 常见连续分布的产生方法

➢$[a,b]$区间上的均匀分布：$X=a+U(b-a)$

➢指数分布：$X=-\ln(U)/\lambda$

➢瑞利分布：$X=\sqrt{-2\sigma^2\ln(U)}$

➢自由度为r的χ^2分布：生成r个独立高斯，求其平方和；

➢对数正态分布：$Y=\exp(X)$, X服从正态分布

- 常见离散分布的产生方法

➢贝努利分布：$B=\begin{cases}1 & U\le p\\ 0 & U>p\end{cases}$

➢二项分布：$X=\sum_{i=1}^{n}B_i$

指数分布：独立随机事件的时间间隔，
泊松分布：单位时间内的随机事件数。

➢泊松分布

- 方法1：产生指数分布随机数并求和，直到大于1为止，输出个数k；
- 方法2：产生均匀分布随机数并求积，直到小于exp(-λ)为止，输出个数k；

➢几何分布：

- 用来表示贝努利实验中，事件第一次发生时所对应的实验次数；
- 概率分布：$P(X=k)=p(1-p)^{k-1}\quad k=1,2,\dots$
- 产生方法：
  - 产生指数分布随机数$Y\sim\exp(\lambda)$，其中λ=-ln(1-p)
  - $X=1+\lfloor Y\rfloor$

- 随机过程的产生
  - 向量舍选法

    > 产生一个满足某种约束条件的向量
    > 方法：产生一个向量，满足约束则接受，不满足则舍去

  - 泊松过程的产生

    泊松过程的两种解释：

    (1)某一时间内的事件发生次数服从泊松分布；

    (2)相邻两次事件发生的时间间隔服从指数分布

    利用第二种解释，可以产生泊松过程

    (1)产生均匀分布随机数

    (2)利用反变换法产生指数分布随机数

    (3)把指数分布随机数作为时间间隔，可以得到一个时间序列

  - 马尔科夫链的产生

    ➢给定初始分布$\pi^{(0)}$和转移矩阵P

    ➢先根据$\pi^{(0)}$产生X_0；

    ➢For t=1:N

    ➢ 根据转移概率矩阵的第X_{t-1}行所对应的概率分布产生X_t;

    ➢最终生成的$X_0\ X_1\ X_2\dots\ X_n$即为马尔科夫链的一个样本。

各种随机量建模

- 非参数建模
  - 生成的随机**输入**为从收集数据中以1/n的概率进行采样得到的数据，其中n为收集数据的总数。
  - **生成随机等概率索引再查找**
  - 建模方式

    1、生成一个随机均匀分布数$u \sim U(0,1)$;

    2、令$P = n \times u$，则$k = [P] + 1$;

    3、 X_k即为生成的随机输入。
- 个体数据的经验建模
  - 将收集到的数据进行递增排列，经验函数为分段线性函数。生成的随机输入x为随机生成的0-1之间的数据在分段函数上对应的值