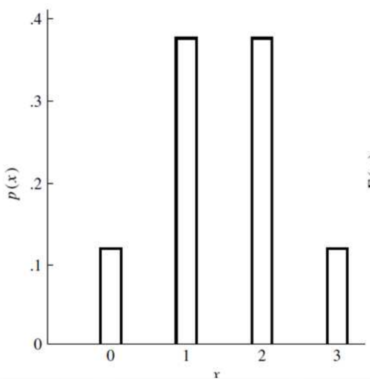
随机变量

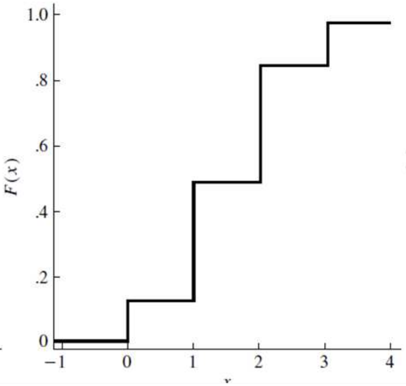
主讲人：刘宏志
liuhz@ss.pku.edu.cn

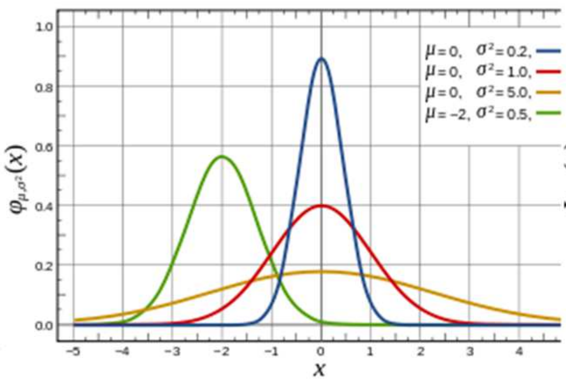
北京大学

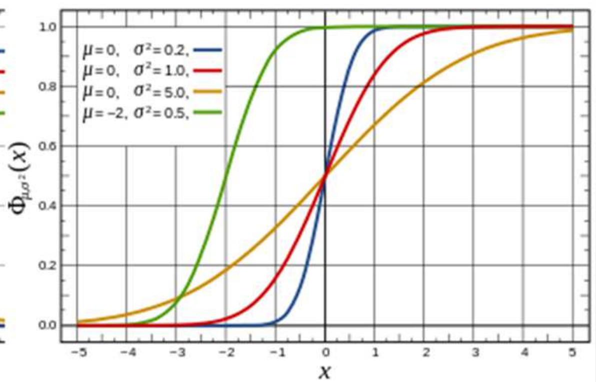
随机变量

- 一个定义在样本空间Ω上的实值函数
  - ➢ 定义了每个样本点$\omega \in \Omega$取值实数的法则
- 本质上是一个随机数
  - ➢ Ω中试验结果的发生是随机的，相应的取值也是随机的
- 概率函数
  - ➢ 质量函数（PMF）：$p(x_i) = P(X{=}x_i)$ （离散型随机变量）
  - ➢ 密度函数（PDF）：$f(x)$ （连续型随机变量）
- 累积分布函数（CDF）：$F(x) = P(X{<=}x)$
  - ➢ 随机变量小于或者等于某个数值的概率$P(X{<=}x)$



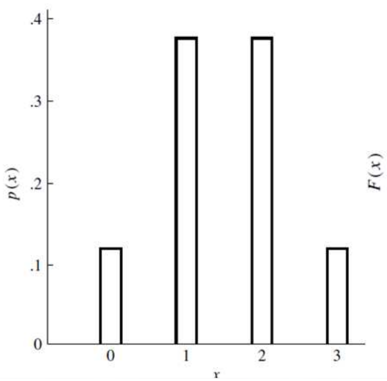


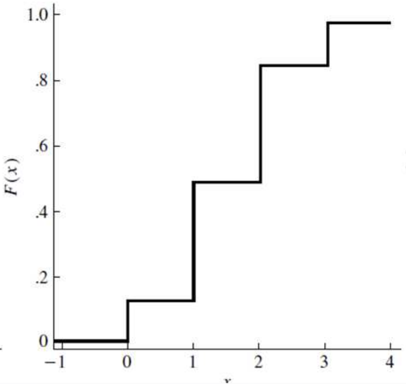


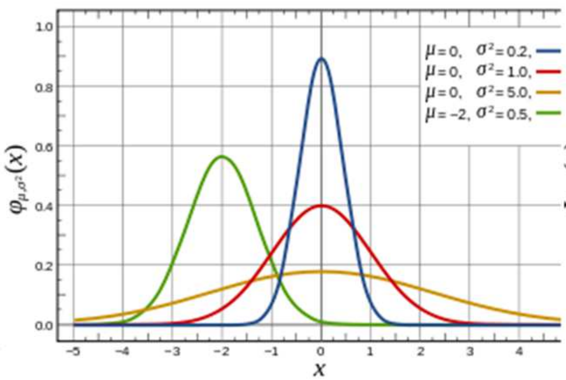


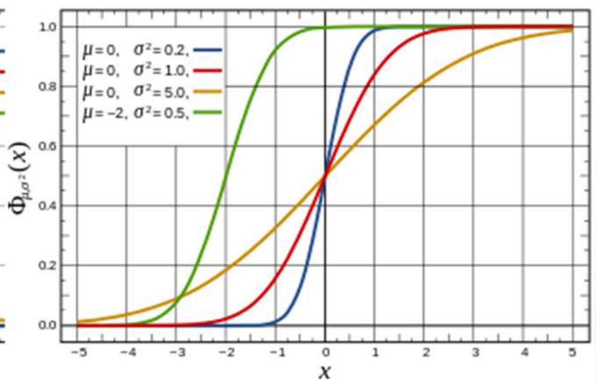
随机变量的类型

- 离散型随机变量：
  - ➢取值是有限或无穷多个
  - ➢例：一次掷20个硬币，x个硬币正面朝上
- 连续型随机变量：
  - ➢分布函数$F(x)$可表示成一个非负可积函数f(x)的积分
  - ➢例：灯泡的寿命x
- 混合型随机变量









随机变量的独立性

- 两个离散型随机变量X和Y是独立的
  - 可能取值分布为$x_1, x_2, \ldots$和$y_1, y_2, \ldots$
  - $\forall\ i, j,\ P(X=x_i,\ Y=y_j) = P(X=x_i)P(Y=y_j)$
- 两个以上离散型随机变量
  - 例如：3个随机变量相互独立
  - $\forall\ i, j, k,\ P(X=x_i,\ Y=y_j,\ Z=z_k) = P(X=x_i)P(Y=y_j)P(Z=z_k)$

常见分布：离散分布

- 伯努利分布
- 二项分布
- 几何分布
- 泊松分布

伯努利分布

- 一种离散型随机分布
- 变量只有两种可能的结果：0和1
  - $p(1)=p, p(0)=1-p$
- 性质：
  - 期望：$E(X)=p$
  - 方差：$D(X)=p(1-p)$

期望：$E(\mathrm{X})=\Sigma x\cdot p(x)$
方差：$D(X)= E\{[x-E(X)]^2\}$

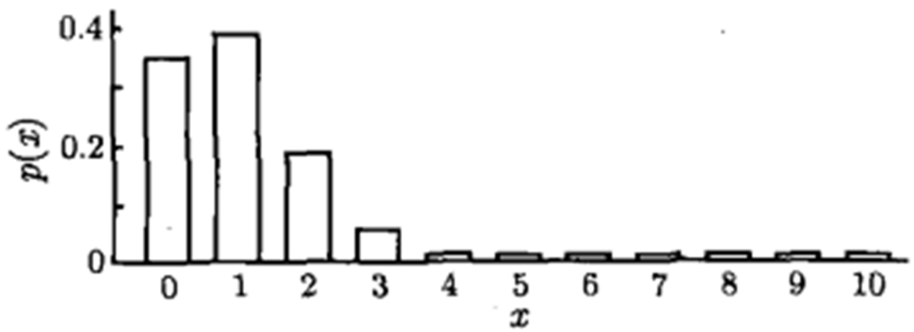
二项分布

- 进行n次独立的试验，每次试验“成功”的概率为p，“失败”的概率为1-p，所有成功的次数X是一个参数为n和p的二项随机变量

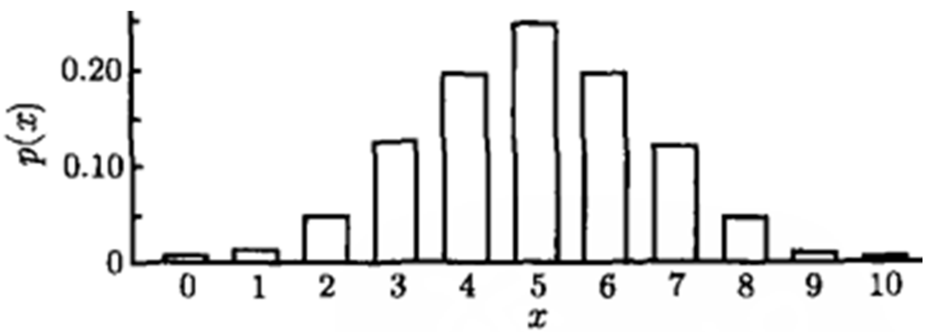
$$p(k) = P(X = k) = C_n^k p^k (1 - p)^{n-k}$$

- 性质：
  - 期望：$E(X)=np$
  - 方差：$D(X)=np(1-p)$

n=1，有什么特性？



a) n=10, p=0.1



b) n=10, p=0.5

二项分布

- 进行*n*次独立的试验，每次试验“成功”的概率为*p*，“失败”的概率为1-*p*，所有成功的次数*X*是一个参数为*n*和*p*的二项随机变量

$$p(k)=P(X=k)=C_n^k p^k (1-p)^{n-k}$$

- 例：掷*n*次硬币的各种结局的出现概率可通过对二项式的*n*次方的展开而得到。*n*=3时

$$(0.5+0.5)^3 = C_3^0 \cdot 0.5^3 + C_3^1 \cdot 0.5^1 \cdot 0.5^2 + C_3^2 \cdot 0.5^2 \cdot 0.5^1 + C_3^3 \cdot 0.5^3$$

$$= 0.125 + 0.375 + 0.375 + 0.125$$

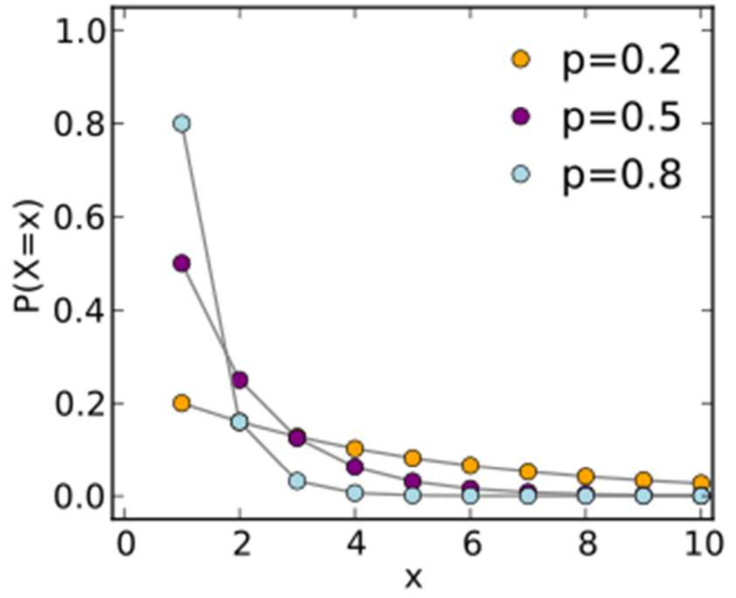
3个正面的概率	2正1反的概率	1正2反的概率	3个反面的概率
0.125	0.375	0.375	0.125

几何分布

- 由无穷次伯努利试验构造而成，每次试验成功的概率为p，X表示直到第一次成功所做的试验次数
  - X=k时，必然前面k-1次试验失败，第k次试验成功

$$p(k)=P(X=k)=(1-p)^{k-1}p$$

- 性质：
  - 期望：$E(X)=1/p$
  - 方差：$D(X)=(1-p)/p^2$

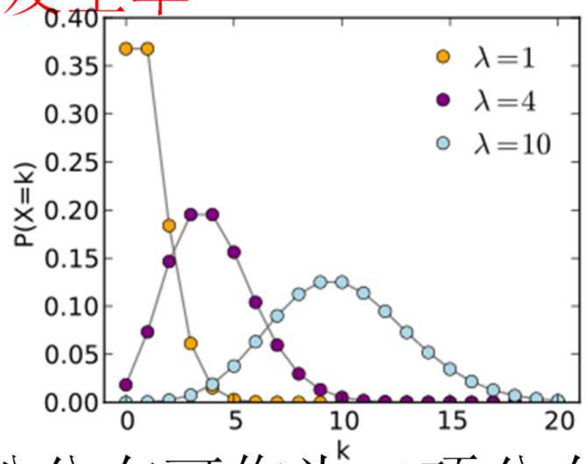


泊松分布

- 单位时间内随机事件发生的次数

$$p(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

  - 参数λ是单位时间内随机事件的平均发生率



- 性质：
  - 期望：$E(X) = \lambda$
  - 方差：$D(X) = \lambda$
- 与二项分布的关系
  - 当二项分布的n很大而p很小时，泊松分布可作为二项分布的近似，其中λ为np
  - 通常当$n \geqq 10, p \leqq 0.1$时，就可用泊松公式进行近似计算

连续随机变量：概率密度函数

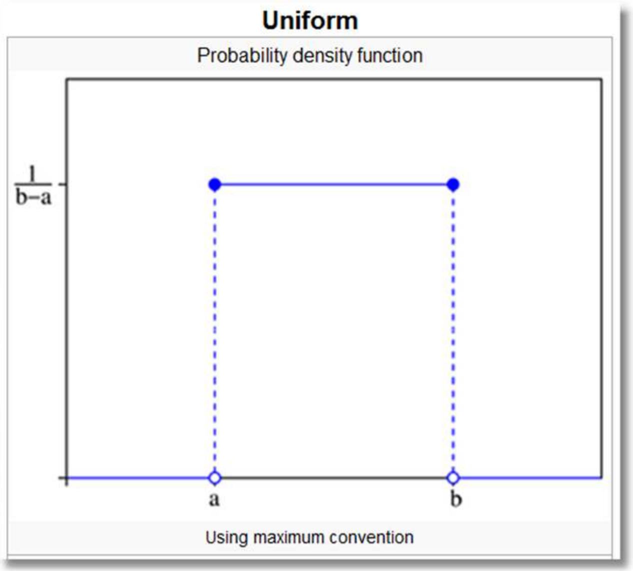
- 性质：
  - $f(x)\geq 0$
  - $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$
- 累积分布函数（CDF）：
  - $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\mathrm{d}u$
- X落在区间(a, b)的概率是$f(x)$从a到b的下方面积：
  - $P(a \leq X \leq b) = \int_{a}^{b} f(u)\mathrm{d}u$
- 连续随机变量X取特定值的概率为0：
  - $P(X=c)=\int_{c}^{c} f(x)\mathrm{d}x=0$

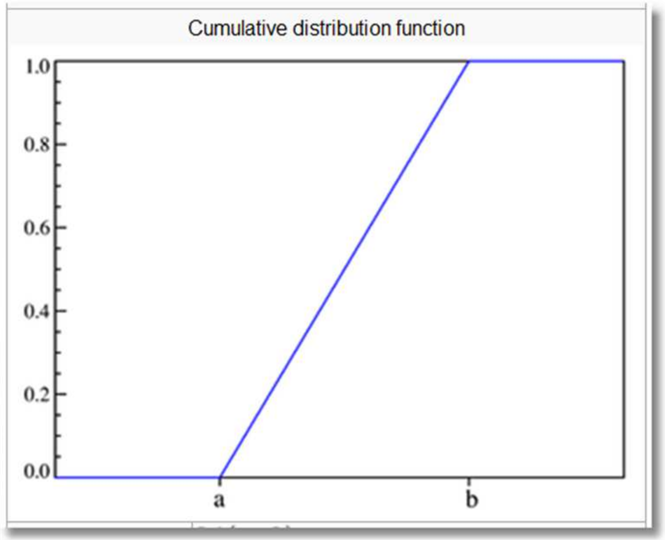
常见分布：连续分布

- 均匀分布
- 指数分布
- 伽马分布
- 正态分布
- 贝塔分布

均匀分布

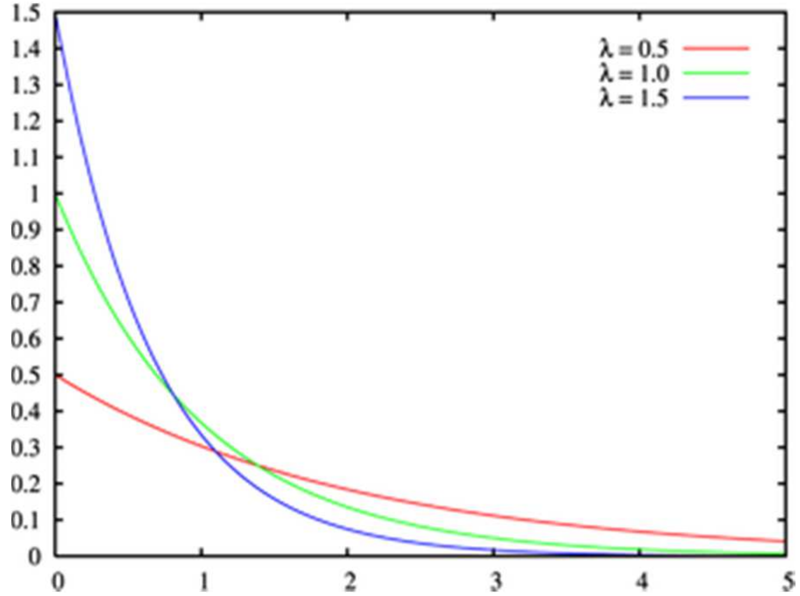
- 概率密度函数：
  - $f(x) = 1/(b-a),\ \ a \le x \le b$
- 累积分布函数：
  - $F(x) = (x\text{-}a)/(b-a),\ a \le x \le b$
- 性质：
  - 期望：E(X) = (a+b)/2
  - 方差：D(X) = $(b\text{-}a)^2/12$





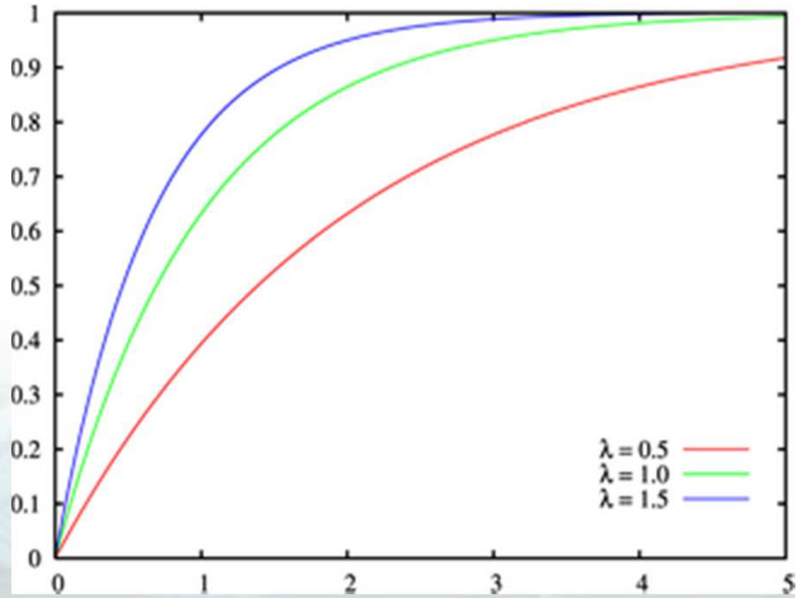
指数分布

- 概率密度函数 $f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0\\ 0\ , & x<0\end{cases}$



- 累积分布函数 $F(x)=\begin{cases}1-e^{-\lambda x}, & x\geq 0\\ 0\ , & x<0\end{cases}$

- 性质：
  - E(X)=1/λ; D(X)=1/λ^2
  - 无记忆性：
    P(X<a+x|X>a)=P(X<x)



指数分布：无记忆性

- P(X<a+x|X>a)=P(X<x)
- 每次事件发生的概率完全不受上次事件发生的影响

$$P(X<a+x\mid X>a)=\frac{P(a<X<a+x)}{P(X>a)}$$

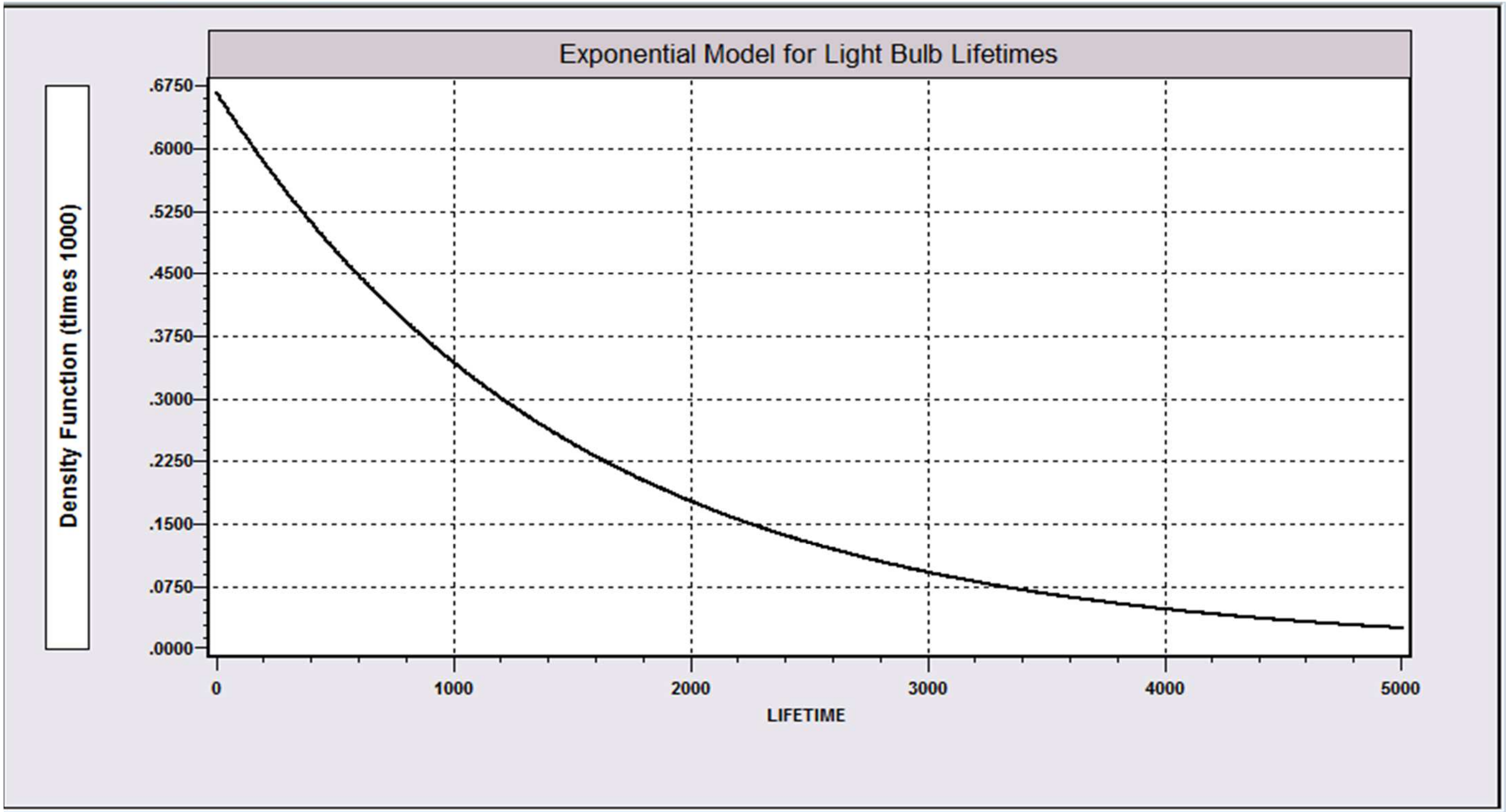
$$=\frac{F(a+x)-F(a)}{1-F(a)}$$

$$=\frac{(1-e^{-\lambda(a+x)})-(1-e^{-\lambda a})}{e^{-\lambda a}}$$

$$=1-e^{-\lambda x}=P(X<x)$$

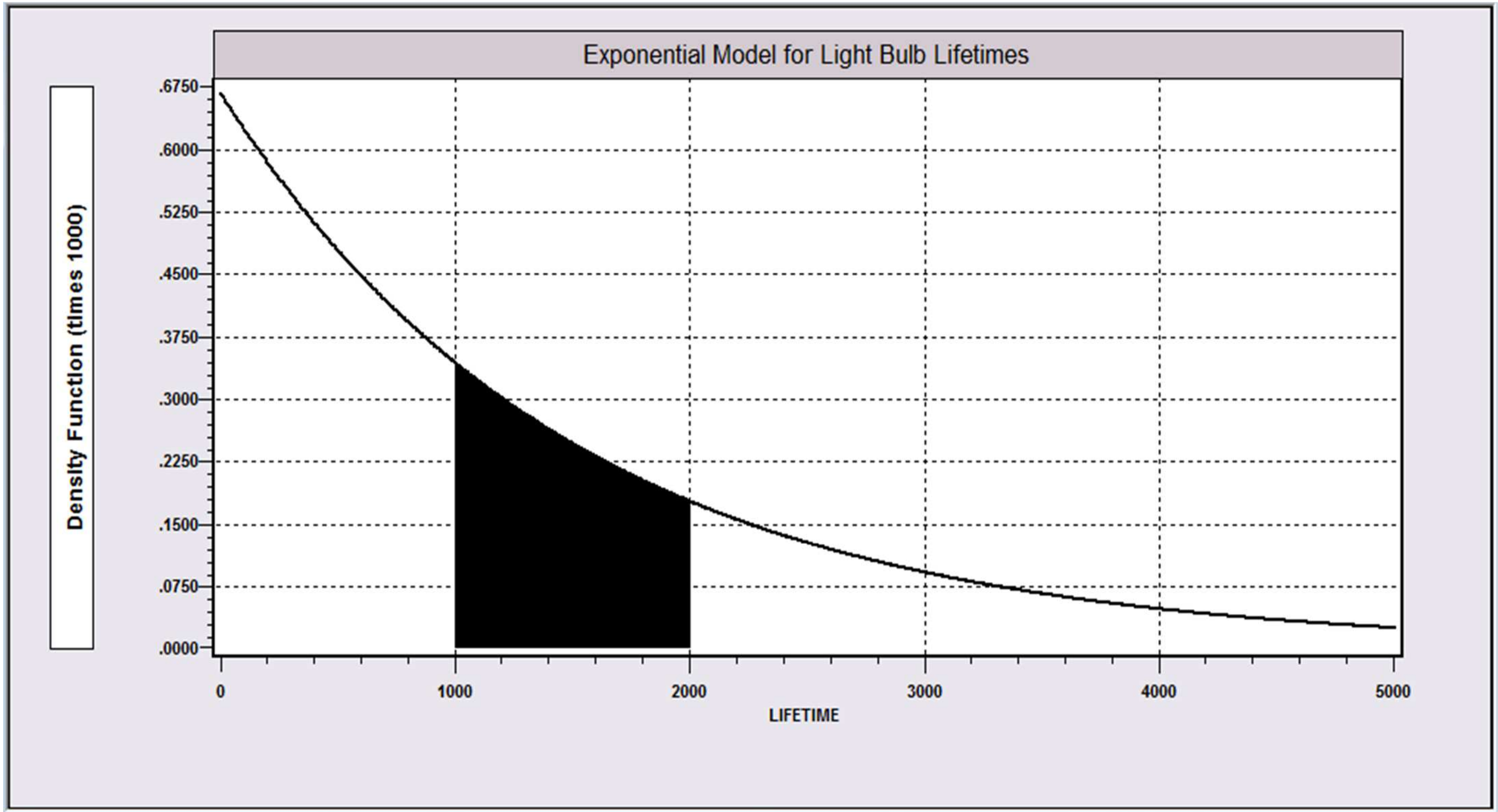
北京大学

案例：灯泡的寿命



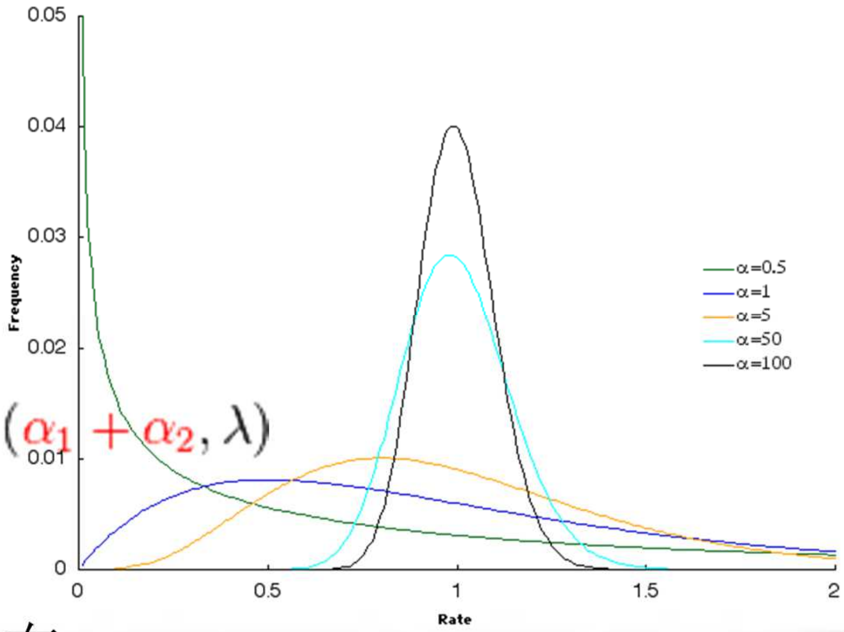
灯泡寿命在1000h~2000h的概率？

案例：灯泡的寿命


Exponential Model for Light Bulb Lifetimes
Density Function (times 1000)
.6750
.6000
.5250
.4500
.3750
.3000
.2250
.1500
.0750
.0000
0
1000
2000
3000
4000
5000
LIFETIME

伽马分布

- 概率密度函数
  - $f(x)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, x\geq 0$
  - 伽马函数：$\Gamma(x)=\int_0^{\infty}u^{x-1}e^{-u}\mathrm{d}u, x>0$
  - α称为形状参数，λ称为尺度参数



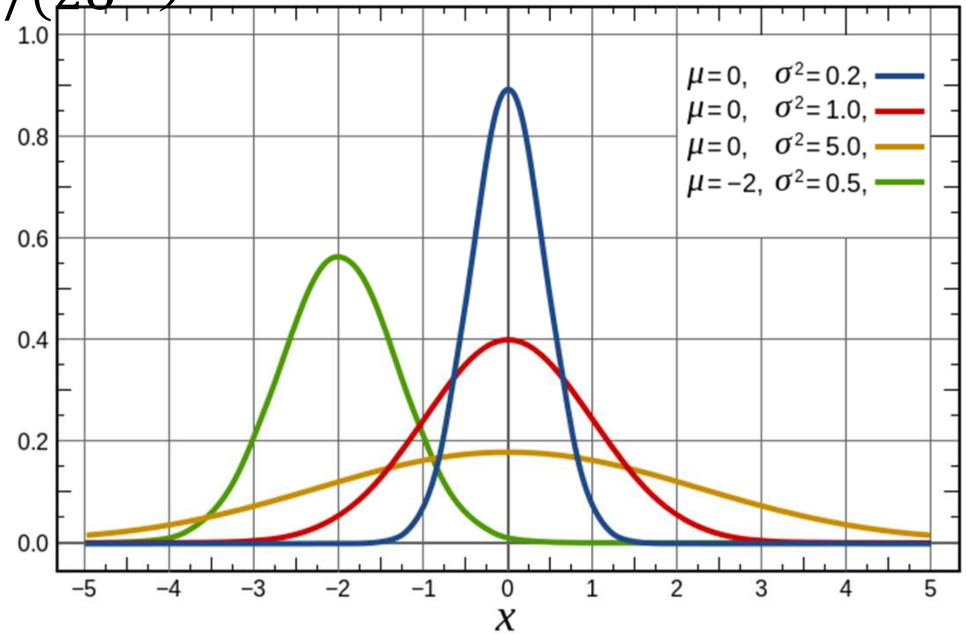
- 性质：
  - 期望：$E(X)=\alpha/\lambda$
  - 方差：$D(X)=\alpha/\lambda^2$
  - 加成性：

$$\mathrm{II}\begin{cases} r.v.X\sim\Gamma(\alpha_1,\lambda) \\ r.v.Y\sim\Gamma(\alpha_2,\lambda)\end{cases} \Longrightarrow X+Y\sim\Gamma(\alpha_1+\alpha_2,\lambda)$$

- 与指数分布的关系：
  - 当α=1时，伽马分布等价于指数分布

正态分布

- 概率密度函数
  - $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$
- 性质：
  - 期望：E(X) =μ
  - 方差：D(X) = σ^2



- 中心极限定理：
  - 许多独立随机变量的和，近似服从正态分布

贝塔分布

- 概率密度函数
  - $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, $0 \leq x \leq 1$
  - 伽马函数：$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} \mathrm{d}u$, $x>0$
  - α和β为形状控制参数
- 性质：
  - 期望：$E(X)=\alpha/(\alpha+\beta)$