概率统计教学讲义

（二）

林文涛编

集美师范专科学校数学科

目　录

第二章 随机变量

第一节 随机变量及其概率分布

§1. 随机变量的概念

在上一章中，我们较详细地讨论了概率论的两个基本概念——随机事件及其概率，使得我们对随机现象的统计规律性有了初步的认识。为了进一步研究随机现象，全面地了解其统计规律性，我们将进一步学习概率论中另外两个重要的基本概念——随机变量及其概率分布。

恩格斯说过："有了变数，运动进入了数学，有了变数，辩证法进入了数学。"那么，我们能不能把随机试验的各种结果——随机事件，也用变数来加以描述呢？实践证明，这不但是完全可能的，而且是十分必要的。其实，在随机现象中，有很大一部份问题就直接与数值发生关系，其结果直接表现为数量，例如"抽查一批产品的次品数""某时刻正在工作的车床数"落入1升水中的细菌数""某产品尺寸的测量误差数"等。有些初看起来与数值无关的随机现象，也常々能联系数值来描述，例如考察某地任一天有否下雨这一随机现象，可能结果共有两个，即"无雨"与"有雨"，与数值没有关系，但是我们可以分别用数"0"与"1"来表示"无雨"与"有雨"，使之与数值发生联系。

一般地，如果A为某个随机事件，则可以通过如下的示性函数使它与数值发生联系：

$$X_A=\begin{cases}1 & \text{如果A发生}\\ 0 & \text{如果A不发生}\end{cases}\qquad (1)$$

可见，随机试验的结果均能用一个数X来表示，而这个数X是随着试验结果的不同而变化的，是样本点的一个函数。我

们把这种随着试验的结果而变化着的量，称为随机变量。一般用大写英文字母X，Y，Z，…，或用希腊字母ξ，η，ζ…来表示。以下给出随机变量的数学定义：

$\boxed{\text{定义}}$：设E是随机试验，E的样本空间是$\Omega=\{\omega\}$，如果对于每一个$\omega\in\Omega$，唯一地对应到一个实数值$X(\omega)$，这样就得到一个定义在Ω上的实值单值函数$X(\omega)$，称$X(\omega)$为随机变量，简记为X。

引入随机变量的概念，就把随机事件数量化了，建立了抽象的样本空间Ω和实数空间R之间的一种对应关系。由于随机变量随试验的结果而取不同的值，所以在试验之前，只能知道它取值的范围，而不能预知它将取什么值。但由于试验的各种结果的出现有一定的概率，所以随机变量取什么值也有一定的概率。这是随机变量与普通函数的本质区别。

$\boxed{\text{例一}}$ 设袋中有5个螺丝钉，3个好，2个坏，从中随机地抽取3个，则"抽得好品"X是一个随机变量，"抽得坏品"Y也是一个随机变量，假定①，②，③为好品，④，⑤为坏品，那么抽取结果ω（样本点）与"好品数"$X(\omega)$、坏品数$Y(\omega)$的对应关系如下表：

样本点 ω	好品数 $X(\omega)$	坏品数 $Y(\omega)$
① ② ③	3	0
① ② ④	2	1
① ② ⑤	2	1
① ③ ④	2	1
① ③ ⑤	2	1
① ④ ⑤	1	2
② ③ ④	2	1
② ③ ⑤	2	1
② ④ ⑤	1	2
③ ④ ⑤	1	2

X只能取1,2,3这三个实值，y只能取0,1,2这三个实值，而$\{X=1\}$ $\{X=2\}$,$\{X=3\}$,$\{y=0\}$,$\{y=1\}$,$\{y=2\}$,这些都是随机事件，而且可求得：

$$P(X=1)=\frac{3}{10} \qquad P(y=0)=\frac{1}{10}$$

$$P(X=2)=\frac{6}{10} \qquad P(y=1)=\frac{6}{10}$$

$$P(X=3)=\frac{1}{10} \qquad P(y=2)=\frac{3}{10}$$

例二　以X表示同安县明年第一代三化螟蛾灯下蛾量，设$\{X=0\}$表示蛾量为"重"；$\{X=1\}$表示蛾量为"中"；$\{X=2\}$表示蛾量为"轻"。X是一个随机变量，可能取值为0,1,2,而且根据以往的统计资料，可以估计出明年灯下蛾量为"重"、"中"、"轻"的概率，既求出$P(X=0)$,$P(X=1)$,$P(X=2)$。

例三　某射手击中目标的概率是p，现在连续向一个目标射击n次，则"击中目标的次数"X是一个随机变量，"第一次击中目标时射击的次数"y也是一个随机变量，X可以取0,1,2,…,n，y可以取1,2,…,n。而且X，y的取值均有一定的概率规律。

以上三个例子的随机变量的取值范围可以一一列举，但实际问题中，有的随机变量的取值不能一一列举，它的取值是连续的，例如"在车站候车的时间"，"气体分子运动的速度"，"弹着点与目标的距离"降水量、洪峰值等々，它们可以在某一区间内，或在$(-\infty,+\infty)$内取值。

例四　厦门轮渡码头每隔5分钟有一辆汽车开往火车站，乘客在任一时刻到达候车站是等可能的。那么，乘客的候车时间X是一个随机变量，显然$\{0\le X<5\}$ 又$\{X>2\}$,$\{X\le4\}$等都是随机事件，而且可以求得它们的概率。

随机变量是概率论与数理统计中一个很重要的概念。因为任何一个随机事件，都可用随机变量的不同取值来表示，这就完全可以把对随机事件的研究转化为对随机变量的研究，而且掌握某一随机变量的变化规律，也就等于了解某一随机现象的整

体性质，使我们对随机现象的认识，从局部而扩大到全局。

实际问题中广泛存在着随机变量，如随机抽取一件产品，那么它的每一项质量指标（如光洁度、硬度、拉力、粘合力等）都是随机变量。要善于把随机变量与我们所要研究的实际问题联系起来。

随机变量一般分为两大类，凡能将其取值一一列举出来（有限或可列）的随机变量称为离散型的随机变量。凡不能将其取值一一列举出来（不可列）的随机变量称为非离散型的随机变量。非离散型的随机变量范围很广，其中最主要的一种是所谓连续型的随机变量（虽不可列，但取值是连续的）。我们只研究离散型随机变量与连续型随机变量。

§2. 离散型随机变量

一、概率分布：

离散型随机变量 X 全部可能的取值是有限个或无穷可列多个，即所有的取值可按一定的顺序排列，表示为数列 $x_1, x_2, \cdots, x_k, \cdots$。

对于随机变量，我们不仅要知道它取值的情况，更主要的是要知道它取各个值的概率，对于离散型随机变量 X，我们除了应该知道它可以取一串值 $X = x_k\ (k = 1, 2, \cdots)$ 以外，更重要的要知道下列一串概率的值：

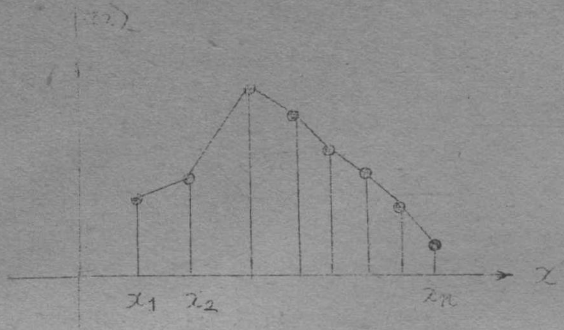
$$p(X=x_1),\quad p(X=x_2),\ \cdots,\ p(X=x_k),\ \cdots$$

我们通常将 X 可能取的值及其相应的概率列成一个表；

X	x_1	x_2	x_3	$\cdots$	x_k	$\cdots$
$p(X=x_k)$	$p(x_1)$	$p(x_2)$	$p(x_3)$	$\cdots$	$p(x_k)$	$\cdots$

称为离散型随机变量的概率分布表，（或称为分布列）有时，为了更加直观，明显地表示离散型随机变量的概率分布，

还可以画出如下的概率分布图



横坐标表示随机变量X的取值，纵坐标表示相应的概率 $p(X=x_k)$ $(k=1,2,\cdots)$ 并用折线将点 $(x_k, p(x_k))$ 连结起来。

概率分布表或概率分布图完整地表示了X的取值及其概率分布情况，全面地描述了离散型随机变量X。有时为了简单起见，也可直接用一系列等式来表示：

$$p_k = p(X=x_k) \qquad (k=1,2,\cdots) \tag{1}$$

这一系列等式就称为X的概率分布，由于事件 $\{X=x_k\}$ $(k=1,2,\cdots)$ 是互不相容的，又是完备的，根据概率的定义，可知分布列 p_k 有以下二条性质。

(1) $p_k \geq 0 \qquad (k=1,2,\cdots)$

(2) $\sum_{k=1}^{\infty} p_k = 1$

二、几种离散型分布：

离散型随机变量的概率分布，就简称为离散型分布，下面介绍几种常见的离散型分布

1.（0—1）分布（二点分布）：若随机变量X只可能取0和1二个值，它的概率分布是

$$p(X=1)=p \qquad (0<p<1)$$
$$p(X=0)=q=1-p \tag{2}$$

则称X服从（0—1）分布（p为参数）。

例一　一批产品有100件，其中有96件正品，现从中随机抽取一件，假定

$$X=\begin{cases}1 & \text{出现正品}\\0 & \text{出现次品}\end{cases}$$

那么有：

$$p(X=1)=96\%$$

$$p(X=0)=4\%$$　　即X服从（0-1）分布.

对于一个随机试验，如果它的样本空间只包含二个元素，即$\Omega=\{\omega_1,\omega_2\}$，我们总能在$\Omega$上定义一个具有（0-1）分布的随机变量：

$$X=X(\omega)=\begin{cases}1 & \text{当}\ \omega=\omega_1\\0 & \text{当}\ \omega=\omega_2\end{cases}$$

然后用它来描述这个随机试验的结果。（0-1）分布虽简单，但经常遇到。

2．二项分布：若随机变量X的概率分布如下：

$$p(X=k)=C_n^k p^k q^{n-k} \qquad \begin{matrix}(k=0,1,2,\cdots,n)\\(0<p<1,\ q=1-p)\end{matrix} \tag{3}$$

则称X服从二项分布（n,p为参数）二项分布常用记号$X\sim B(n,p)$来表示。容易证明，二项分布满足分布列的二条性质：

(1)　$p(X=k)\geq 0\quad(k=0,1,2,\cdots,n)$

(2)　$\sum\limits_{k=0}^{n}p(X=k)=\sum\limits_{k=0}^{n}C_n^k p^k\cdot q^{n-k}=(p+q)^n=1$

二项分布见于贝努里概率，设事设A发生的概率为p，以X表示n重贝努里试验中事件A发生的次数。那么，事件A恰发生k次的概率$P_n(k)$服从二项分布，即

$$P_n(k)=p(X=k)=C_n^k p^k\cdot q^{n-k}\qquad k=0,1,2,\cdots,n$$

容易看出，二项分布当$n=1$时，就是（0-1）分布，即

单次试验的情况。

3. 普阿松（paisson）分布：若随机变量X的概率分布如下：

$$p(X=k)=\frac{\lambda^k}{k!}e^{-\lambda} \quad (k=0,1,2,\cdots;\ \lambda>0) \quad (4)$$

则称X服从普阿松分布（λ为参数）。

注意到 $e^x=1+\frac{x}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\cdots$，因此容易证明：

$$\sum_{k=0}^{\infty}p(X=k)=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}=e^{-\lambda}\cdot\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}\cdot e^{\lambda}=1$$

普阿松分布在历史上是作为二次分布的近似分布，於1837年由法国数学家普阿松引入的。许多随机现象服从普阿松分布，例如电话交换台在某一时刻来到的呼叫次数，公共汽车站来到的乘客数，放射性物质落到某区域的质点数，显微镜下落在某区域中的微生物数等々都服从普阿松分布。

例二 放射性物质在某一段时间内放射出的粒子数X服从普阿松分布。1910年罗塞福（Rutherford）和格吉（Geiger）做了一个著名的试验，观察放射性物质放出α粒子个数的情况，做了2608次观察，每次观察时间是7.5秒。记录如下表：

放射粒子数 X	观察到次数 m_k	频率 $W(X=k)=\frac{m_k}{N}$	按普阿松分布计算概率 $p(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$
0	57	0.022	0.021
1	203	0.078	0.081
2	383	0.147	0.156
3	525	0.201	0.201
4	532	0.204	0.195
5	408	0.156	0.151
6	273	0.105	0.097
7	139	0.053	0.054
8	45	0.017	0.026
9	27	0.010	0.011
≥10	16	0.006	0.007
总计	2608	0.999	1.000

从表中可看出按（4）式算出的概率 $p(X=k)$ 跟 $\{X=k\}$ 的频率相当接近。

其中：$\lambda=\frac{10094}{2608}=3.87$，10094 为 2608 次放出的 α 粒子总数，λ 表示平均每次放出的 α 粒子数。

我们还可以从理论上来加以分析推导，说明为什么放射的粒子数 X 服从普阿松分布。设想将体积为 V 的放射性物质分为体积相同的 n 个小块，每小块体积 $\Delta V=\frac{V}{n}$，并假定：

(1) 每一小块在一定时间内（例如 7.5 秒里）放出二个以上 α 粒子的概率为 0（其实是概率很少，可忽略不计），而放出一个 α 粒子的概率为 p_n，p_n 只跟体积 ΔV 有关。

$$p_n=\mu\Delta V \quad (\mu \text{为比例系数})$$

(2) 每一小块 ΔV 是否放出 α 粒子，是相互独立的。

在这两条假定下，体积为 V 的放射性物质，在一定的时间内放出的 α 粒子数 k，可以看成在 n 个独立的小块中，恰有 k 块放出 α 粒子，于是，放出 k 个粒子的概率，就可按独立试验序列来近似计算：

$$p(X=k)\approx C_n^k p_n^k\cdot q_n^{n-k} \qquad (q_n=1-p_n)$$

当 $n\to\infty$ 时，即将 V 无限细分时，近似就转化为精确。

$$p(X=k)=\lim_{n\to\infty} C_n^k p_n^k\cdot q_n^{n-k}$$

$\because$ $p_n=\mu\Delta V=\frac{\mu V}{n}$，记 $\lambda=\mu V$，那么有 $\lambda=np_n$，$p_n=\frac{\lambda}{n}$。

将 $p_n=\frac{\lambda}{n}$ 代入，可以证明：

$$p(X=k)=\lim_{n\to\infty} C_n^k\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}=\frac{\lambda^k}{k!}e^{-\lambda} \quad (k=0,1,2,\cdots)$$

这就是第一章已证明过的普阿松定理（详细证明见第一章），说明了普阿松分布是二项分布，当 $np=\lambda$，$n\to\infty$ 情况下的极限分布，所以，当 n 很大，p 很小时，可用普阿松分布作为二

项分布的近似计算。

具体问题中的随机现象，只要它符合类似于(1)、(2)的假定，那么，就会出现服从普阿松分布的随机变量。

4．超几何分布：若随机变量X的概率分布如下：

$$p(X=k)=\frac{C_n^k\,C_{N-M}^{n-k}}{C_N^n}\qquad \begin{pmatrix} k=0,1,2,\cdots,l \\ l=\min(M,N) \end{pmatrix}\qquad (5)$$

则称X服从超几何分布。

N件产品中有M件次品，从中随机抽取n件，（假定$n\leq N-M$）则M件中含有的次品数X服从超几何分布。

现在我们来证明，服从超几何分布的随机变量X，当$N\to\infty$，$\frac{M}{N}\to p$时近似地服从二项分布．即证明

$$\lim_{\substack{N\to\infty\\ \frac{M}{N}\to p}}\frac{C^k\cdot C_{N-M}^{n-k}}{C_N^n}=C_n^k\cdot p^k\cdot q^{n-k}$$

证明：$p(X=k)=\dfrac{C_M^k\cdot C_{N-M}^{n-k}}{C_N^n}=\dfrac{M!}{k!(M-k)!}\cdot\dfrac{(N-M)!}{(n-k)!\{N-M-(n-k)\}!}\cdot\dfrac{n!(N-n)!}{N!}$

$$=\frac{n!}{k!(n-k)!}\cdot\left\{\frac{M\cdot(M-1)\cdots(M-k+1)}{\underbrace{N\cdot N\cdots N}_{k个}}\right\}\cdot$$

$$\cdot\left\{\frac{(N-M)(N-M-1)\cdots\{N-M-(n-k)+1\}}{\underbrace{N\cdot N\cdot N\cdots N}_{n-k}}\right\}\cdot$$

$$\cdot\left\{\frac{\overbrace{N\cdot N\cdots\cdot N}^{n个}}{N(N-1)\cdots(N-n+1)}\right\}$$

∵ 当$N\to\infty$，$\frac{M}{N}\to p$时，第一个括号$\to p^k$；第二个括号$\to q^{n-k}$；第三个括号$\to 1$，

∴ 此时，$p(X=k)\longrightarrow C_n^k\,p^k\,q^{n-k}$，即

$$\lim_{\substack{N\to\infty\\ \frac{M}{N}\to p}}\frac{C_M^k\cdot C_{N-M}^{n-k}}{C_N^n}=C_n^k\cdot p^k\cdot q^{n-k}$$

5. 几何分布：若随机变量X的概率分布如下：

$$p(X=k)=q^{k-1}\cdot p \qquad k=1,2,\cdots \tag{6}$$

则称X服从几何分布。在贝努里试验中，等待事件A首次成功的次数X服从几何分布。

例三　某射手连续向同一目标射击，击中目标的概率为0.8，那么第一次击中目标时共射击了k次的概率为：

$$p(X=k)=0.2^{k-1}\times 0.8 \qquad (k=1,2,\cdots)$$

6. 巴斯卡分布：若随机变量X的概率分布如下：

$$p(X=k)=C_{k-1}^{r-1}\ q^{k-r}\cdot p^{r} \qquad (k=r+1,r+2\cdots) \tag{7}$$

则称X服从巴斯卡分布。在贝努里试验中，第r次成功所需要的次数X服从巴斯卡分布。

例四　某射手连续向同一目标射击，击中目标的概率为p，击不中目标的概率为q（$p+q=1$），求第r次击中目标所需要次数的概率分布。

解：设在k次射击中，r次击中，k－r次击不中。例如前k－r次击不中，後r次击中，那么概率为$q^{k-r}\cdot p^{r}$。根据题意最後一次是一定要击中的，那么前面的（k－1）次中，（k－r）次击不中的情况，相当於从（k－1）个元素中取（k－r）个元素的组合数，所以r次击中，（k－r）次击不中的所有可能情况有$C_{k-1}^{k-r}=C_{k-1}^{r-1}$。因此可得到：

$$p(X=k)=C_{k-1}^{r-1}\ q^{k-r}p^{r} \qquad (k=r+1,r+2\cdots)$$

7. 退化分布（单点分布）：若随机变量X只取一个常数c，即

$$p(X=c)=1$$

则称X服从退化分布。其实，退化分布并不随机，如一年级一班学生的年龄都是七岁（记为X），所以有 $p(X=7)=1$。正因为如此，才称为退化分布，只是为了统一，才称它为随机变量 $\{X=7\}$ 为必然事件，$\{X\neq 7\}$ 为不可能事件。

§3. 连续型随机变量

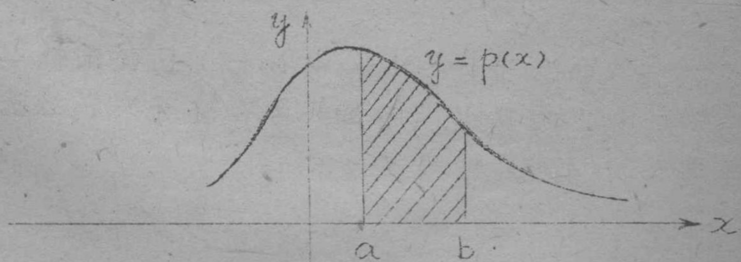
一、概率密度函数：

连续型随机变量的特点是可以取某一区间内的任何数值，由于这些数值构成不可数的无穷集合。因此，我们不可能用描述离散型随机变量的办法来描述它。对于非离散型的随机变量，考察事件 $\{X=a\}$ 发生的概率往往意义不大，虽然 $\{X=a\}$ 决不是不可能事件，但我们只能认为 $p(X=a)=0$。例如，测量某一零件尺寸时，我们只能说测量偏差为+0.030001 mm的概率等于零。实际上，我们也不可能去检验事件 $\{X=+0.003001\}$ 是否发生。因此，对于连续型随机变量，我们往往考察X落在某一区间内的概率。为此引出如下关于概率密度函数的定义。

$\boxed{\text{定义}}$：对于随机变量X，若存在非负可积函数 $p(x)$ $(-\infty<x<+\infty)$，使对任意 $a<b$，成立：

$$p\{a<X<b\}=\int_a^b p(x)\,dx \qquad (1)$$

则称X为连续型随意变量。$p(x)$ 称为连续型随机变量X的概率分布密度函数，简称分布密度或密度函数。$p(x)$ 的图象称为X的分布曲线，曲线 $p(x)$ 与x轴介于a、b之间的曲边梯形的面积表示事件 $\{a<x<b\}$ 的概率。



关于概率密度函数 $p(x)$，也有以下二条性质：

(1) $p(x) \geq 0$

(2) $\int_{-\infty}^{+\infty} p(x)dx = p(-\infty < x < +\infty) = p(\Omega) = 1$.

概率密度的"密度"一词，跟物理学中质量线密度的"密度"有相似之处。若 $p(x)$ 在 $x = x_0$ 连续时，由定积分的性质得：

$$\lim_{\Delta x \to 0} \frac{p(x_0 < x < x_0 + \Delta x)}{\Delta x} = p(x_0) \quad \text{-------- (2)}$$

由此可见，当 $p(x)$ 大时，X 在 x_0 附近取值的概率也就较大，或者说"密度"较大。

二. 几种连续型分布：

1. 均匀分布（等概率分布）：若随机变量 X 的概率分布密度函数为：

$$p(x) = \begin{cases} \lambda & \text{当 } a \leq x \leq b \\ 0 & \text{当 } x < a \text{ 或 } x > b \end{cases} \qquad (3)$$

则称 X 服从 $[a, b]$ 区间上的均匀分布。

设 X 为 $[a, b]$ 上的均匀分布，由于

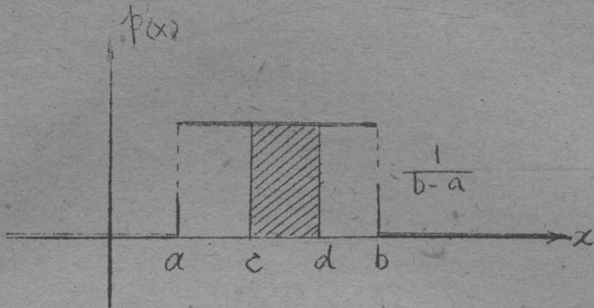
$$p(-\infty < x < +\infty) = p(a < x < b) = \int_a^b \lambda dx = \lambda x \Big|_a^b =$$

$$= \lambda(b - a) = 1$$

所以有 $\lambda = \frac{1}{b-a}$，并且对于任意满足 $a \leq c < d \leq b$ 的 c, d 有

$$p(c < x < d) = \int_c^d p(x)dx = \lambda(d - c) = \frac{d-c}{b-a}.$$

由此可见，X 取值于 $[a, b]$ 中任一小区间的概率与该小区间的长度成正比，而与小区间的具体位置无关，这就是均匀分布的概率意义。$p(x)$ 的图象如下：



在数值计算中，四舍五入所引起的误差X，可以看作是一个服从$[-0.5\times10^{-n}, 0.5\times10^{-n}]$上的均匀分布。（保留n位小数）

§1. 例四的候车时间X服从[0, 5]上的均匀分布.

2. 指数分布：若随机变量X的密度函数为

$$p(x)=\begin{cases}\lambda e^{-\lambda x} & \text{当 } x\geq 0\\ 0 & \text{当 } x<0\end{cases} \qquad (\lambda>0 \text{ 为常数}) \qquad (4)$$

则称X服从参数为λ的指数分布.

指数分布可近似地表达各种"寿命"的分布，例如无线电元件的寿命，动物的寿命，电话问题中的通话时间，随机服务系统中的服务时间等。

若X服从参数为λ的指数分布，对任何 $0\leq a<b$，有

$$p(a<X<b)=\int_a^b \lambda e^{-\lambda x}dx=\int_{\lambda a}^{\lambda b} e^{-t}dt=-e^{-t}\Big|_{\lambda a}^{\lambda b}=$$

$$=e^{-\lambda a}-e^{-\lambda b}$$

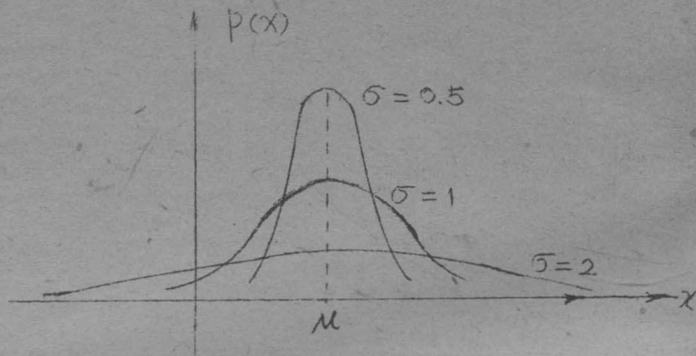
因此可得 $\int_{-\infty}^{+\infty}\lambda e^{-\lambda x}dx=\int_0^{+\infty}\lambda e^{-\lambda x}dx=e^0-e^{-\infty}=1$

3. 正态分布（高斯分布）：若随机变量X的密度函数为

$$p(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \qquad (-\infty<x<+\infty) \qquad (5)$$

其中 $\sigma>0$，μ和σ均为常数，则称X服从正态分布，简记为 $N(\mu,\sigma^2)$。特别地，当 $\mu=0$，$\sigma=1$，即 $p(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ 时，称为标准的正态分布，简记为N(0, 1)。习惯上把服从正态分布的随机变量称为正态变量。

一般正态分布密度函数 $p(x)$ 的图形如下：



$p(x)$在 $x=\mu$ 处达到极大值，等于$\frac{1}{\sqrt{2\pi}\sigma}$；曲线相对于直线 $x=\mu$ 对称；在 $x=\mu\pm\sigma$ 处有拐点；当 $x\to\pm\infty$ 时，曲线以 x 轴为其渐近线；当 σ 不同时，$p(x)$ 的形状也不同．σ 越小，分布越集中．在 $x=\mu$ 附近，曲线陡峭；当 σ 越大时，分布越平坦，曲线平缓．这是因为分布曲线 $p(x)$ 下面的面积总保持等于 1，所以 σ 值越小时，曲线在中心部份的纵坐标增大的缘故。

我们可以验证：$p(-\infty<x<+\infty)=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=1$

证明：令 $\frac{x-\mu}{\sigma}=y$，则

$$\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy$$

又 $$\left(\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx\right)\left(\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy\right)=$$

$$=\frac{1}{2\sigma}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{x^2+y^2}{2}}dxdy$$

作极坐标变换，令 $x=r\cdot\cos\varphi$，$y=r\cdot\sin\varphi$，则

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{x^2+y^2}{2}}dxdy=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{r^2}{2}}r\cdot drd\varphi$$

$$=\frac{1}{2\pi}\int_{0}^{+\infty}e^{-\frac{r^2}{2}}\cdot r\left[\int_{0}^{2\pi}d\varphi\right]dr=\int_{0}^{\infty}e^{-\frac{r^2}{2}}rdr=\int_{0}^{\infty}e^{-u}du=1$$