思考：

1. 在实际中如何判断一个随机变量是连续型随机变量？

2. 如何通过一个分布函数来判断随机变量的类型？

本次课的重点是：

连续型随机变量的密度函数与分布函数的判定及性质，均匀分布及相关的计算问题.

下次课内容：

指数分布及正态分布的密度函数与分布函数的性质、计算及应用，二维随机变量联合分布性质.


4:24
关注
朋友
推荐
人工智能霸榜
诺贝尔物理学奖和化学奖
Google DeepMind 员工
靠蛋白质结构预测获奖
NOBELPRISET I KEMI 2024
THE NOBEL PRIZE IN CHEMISTRY 2024
KUNGL. VETENSKAPS AKADEMIEN
David Baker
University of Washington
USA
Demis Hassabis
Google DeepMind
United Kingdom
John M. Jumper
Google DeepMind
United Kingdom
"för datorbaserad proteindesign"
"for computational protein design"
"för proteinstrukturprediktion"
"for protein structure prediction"
#NobelPrize
Announcement of the 2024 Nobel Prize in Chemistry
授予大卫·贝克（David Baker）
欢迎关注 新智元
诺贝尔物理学奖和化学奖，
都颁给了搞AI的了…
新智元
1个朋友关注
+关注
956
1771
596
111


解读：物理诺贝尔奖为何颁给了 HNN 之父和深度学习之父?
原创 刘洁、郑佳美 AI科技评论 2024年10月08日 19:58 广东
THE NOBEL PRIZE
IN PHYSICS 2024
Illustrations: Niklas Elmehed
John J. Hopfield
Geoffrey E. Hinton
"for foundational discoveries and inventions
that enable machine learning"


11:14
诺贝尔化学奖出炉
诺贝尔化学奖出炉
澎湃新闻
2024-10-09 17:49 北京
+关注
「原标题：2024年诺贝尔化学奖揭晓，三位科学家获殊荣」
当地时间10月9日，瑞典皇家科学院宣布，将2024年诺贝尔化学奖授予三位科学家，其中，一半授予美国华盛顿大学的大卫·贝克（David Baker），以表彰其在计算蛋白质设计方面的贡献，另一半则共同授予Demis Hassabis和John M. Jumper，以表彰其在蛋白质结构预测方面的贡献。
THE NOBEL PRIZE
IN CHEMISTRY 2024
Illustrations: Niklas Elmehed
App内打开
写跟贴
1.1万
1109

研究一个随机变量 (random variable)
- (1)首先关注取值情况，即取哪些值.
- (2)更关注它取各种值的概率情况，即“概率的分布”.

离散型随机变量
- 只可能取有限个或可列无穷多个值.
- 分布律(最方便的标志), 分布函数(阶梯).

连续型随机变量
- 充满某个区间，无法一一列出.
- 密度函数(最方便的标志), 分布函数(连续).

任何类型随机变量
- $P\{X=x\}=F(x)-F(x-0), \forall x\in(-\infty,+\infty)$
- $P\{a<X\le b\}=F(b)-F(a)$

由于在许多实际问题中出现二项分布,并且计算显得非常重要.
法国数学家泊松在1837年引入泊松分布，作为二项分布的近似.

一、二项分布

记为$X \sim B(n, P(A))$

在n重贝努里试验中,事件A(关注试验两个结果中的一个)发生的次数X的分布律$P\{X = k\} = C_n^k p^k (1-p)^{n-k}$, $k=0,1,2,\dots,n$

负二项分布: 事件A发生k次时的总试验次数Y或者对立事件发生的次数W?;

$P\{Y = t\} = C_{t-1}^{k-1} p^k q^{t-k}, t = k, k+1, \dots$

$P\{W = w\} = C_{w+k-1}^{k-1} p^k (1-p)^w, w = 0,1,2,\dots.$

二、泊松分布

记为 $Y \sim P(\lambda)$

在Y表示在一定的时间或空间内出现的事件个数

分布律为$P\{Y=k\}= \dfrac{\lambda^k}{k!} e^{-\lambda}$, $k=0, 1, \dots, \lambda > 0.$

联系: 一定条件下泊松分布可视为二项分布的极限分布

若 $\lim\limits_{n\to\infty} np = \lambda > 0$,(实际问题)当$n$够大,$p$较小时(稀有事件),

$$P(X= k) = C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda} \quad k=0,1,2,\dots,n$$

连续型随机变量的概率密度(标志):

（1）$f(x) \geqslant 0$ (2) $\int_{-\infty}^{+\infty} f(x)dx = 1$(归一性)

概率密度与分布函数的关系:

$f(x) \Rightarrow F(x)$ $F(x) = \int_{-\infty}^{x} f(t)\,dt$

$F(x) \Rightarrow f(x)$ $f(x) = \begin{cases} F'(x), & F'(x)\text{存在的区间} \\ 0 & else \end{cases}$

相关的性质: 或者 $f(x) = F'(x),\ x \in (-\infty, +\infty)$

(1) $F(x) = F(x+0) = F(x-0)$ (2) $x_0 \in R,\ P\{X = x_0\} = 0.$

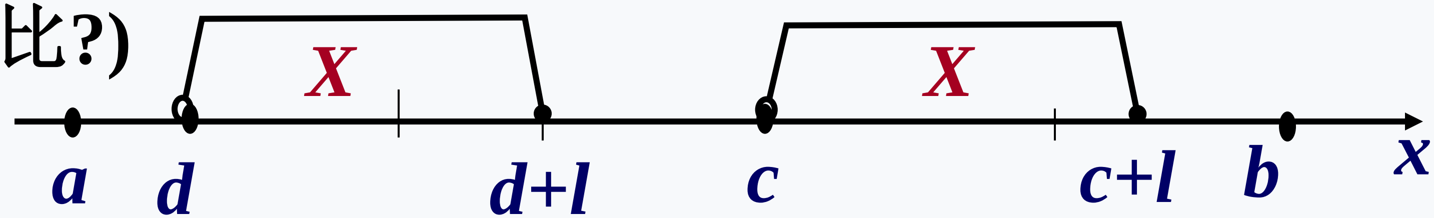
(3) $P\{x_1 < X \le x_2\} = P\{x_1 \le X \le x_2\} = \int_{x_1}^{x_2} f(x)dx$

（1）均匀分布：$X \sim U(a, b)$

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a<x<b\\ 0 & \text{其它}\end{cases}$$

密度函数：概率在各处的"密集程度"一样，即概率均匀分布在这区间上.

特点：随机变量落在(a, b)的子区间的概率与位置无关，仅与测度(即长度)成正比((几何概率)样本点是等可能的：概率与位置，形状 均无关，与测度(长度、面积、体积)成正比?)



$$\frac{l}{b-a}$$

$$X \sim U(a,b),\ F_X(x)=\begin{cases}0, & x\le a,\\ \dfrac{x-a}{b-a}, & a<x\le b,\\ 1, & x>b.\end{cases}$$

$$Y\sim U(0,1),\ \ F_Y(y)=\begin{cases}0, & y\le 0,\\ y, & 0<y<1,\\ 1, & y\ge 1\end{cases}$$

设随机变量X的分布函数$F(x)$连续且单调增加，求证：随机变量$Y=F(X)\sim U(0,1)$.

证明：由于$F(\cdot)\uparrow$，$F^{-1}(\cdot)$存在，$Y=F(X):\Omega\to[0,1]$

当$0<y<1$时，$F_Y(y)=P\{Y\le y\}=P\{F(X)\le y\}$

$$=P\{X\le \underline{F^{-1}(y)}_{\,x}\}$$

$$=F_X(F_X^{-1}(y))=y$$

举例：$F(x)=x^2$,

则 $F(X)=X^2$

$$f_Y(y)=\begin{cases}1, & 0<y<1\\ 0, & else.\end{cases}$$

或者 $F_Y(y)=\begin{cases}0, & y\le 0,\\ y, & 0<y<1,\\ 1, & y\ge 1\end{cases}$

应用：随机模拟技术中，服从$U(0,1)$的随机变量是最基本的一类，是其它随机变量的计算机模拟的基础.

4. 如果随机变量 Y 的分布函数 F(y)连续且严格单调增加，而随机变量 $X\sim U(0,1)$ 令 $Z=F^{-1}(X)$，那么 Z 与 Y 同分布，请说明理由。

4. 理由：因为$X\sim U(0,1)$，所以

$$F_X(x)=\begin{cases}0, & x\le 0\\ x, & 0<x\le 1\\ 1, & x>1\end{cases}$$

所以：

$$\begin{aligned}F_Z(y)&=P\{Z\le y\}\\&=P\{F^{-1}(X)\le y\}\\&=P\{X\le \underline{F(y)}\}\\&=F_X(F(y))\\&=F(y)\end{aligned}$$

x

应用: 随机模拟技术中, 服从$U(0,1)$的随机变量是最基本的一类, 利用以上例子可产生服从各种分布的随机变量供使用.

指数分布的由来：使用了t小时的电子管在以后的Δt小时内损坏的概率等于$\lambda\Delta t + o(\Delta t)$，其中$\lambda > 0$为一常数(失效率或故障率)，试写出电子管的寿命$T$的分布函数.

解：$P\{T \le t + \Delta t \mid T > t\} = \lambda\Delta t + o(\Delta t), \quad \Delta t > 0$

$$\lambda = \lim_{\Delta t \to 0+} \frac{P\{T \le t + \Delta t \mid T > t\}}{\Delta t}$$

失效率：元件在t时刻仍正常,在将来单位长度时间内失效或故障的概率.

求解方程得分布函数

$$F(t) = \begin{cases} 1 - e^{-\lambda t}, & t \ge 0; \\ 0, & t < 0. \end{cases}$$

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & t \ge 0; \\ 0, & t < 0. \end{cases}$$

结论：由失效率λ推导出指数分布.

三、指数分布

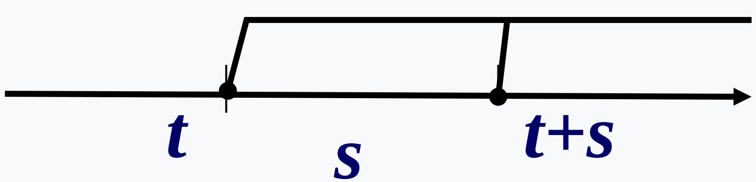
设随机变量X的概率密度函数为

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases}\qquad(\lambda>0)$$

则称随机变量X服从参数为λ的指数分布.

特点：指数分布具有无记忆性, 即有(教材例2.3.4).

$$P\{X>t+s\mid X>t\}=P\{X>s\}$$



永远年青分布

注: 如果已知某人的年龄为s, 则再活t年的概率与年龄s无关, 所以有时又风趣地称指数分布“永远年轻”的.

例5.某电子元件发生故障则不可修复，它的寿命X服从参数为λ=1/2000的指数分布，它工作了1000小时后能再工作1000小时的概率为多少？

解：$P\{X \geq 2000 \mid X \geq 1000\} = P\{X \geq 1000\}$

$=1-P\{X<1000\}=1-F(1000)$

$=1-[1-e^{-1000/2000}]=e^{-1/2}$.

其中

$$F(x)=\begin{cases} 1-e^{-\lambda x}, & x>0; \\ 0, & x \leq 0 \end{cases}$$

性质：$P\{X>t+s \mid X>t\}=P\{X>s\}$(无记忆性)

三、指数分布

指数分布与泊松分布的关系?

设随机变量X 的概率密度函数为

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases}\quad(\lambda>0)$$

则称随机变量X 服从参数为 λ 的指数分布.

应用: 指数分布常用于描述处于稳定工作状态的电子元件的寿命, 电话问题中的通话时间, 随机服务系统中的服务时间等.

在电子元器件的可靠性研究中, 在日本的工业标准和美国军用标准中:

大型复杂系统(如计算机)的故障间隔时间的失效分布都是采用指数分布描述.(思考为什么?)

三、指数分布

$$F(t)=\begin{cases}1-e^{-\lambda t}, & t\ge 0;\\ 0, & t<0.\end{cases}$$

设随机变量X的概率密度函数为

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases}\quad(\lambda>0)$$

则称随机变量X服从参数为λ的指数分布.

指数分布与泊松分布的关系?

$X(t)$(随机变量) ~ $P(\lambda_0 t)$:书上等给出泊松分布的参数$\lambda=\lambda_0 t$

参数λ_0:单位时间$\lambda_0 t$内事件出现的平均次数(平均到达率或强度).

若次数$X(t)$为参数为λt的泊松分布, 以σ_1记为第一次到达的时刻, 则

$$P\{\sigma_1\ge t\}=P\{X(t)=0\}=\frac{(\lambda t)^0}{0!}e^{-\lambda t},t\ge 0$$

$$F_{\sigma_1}(t)=P\{\sigma_1\le t\}=1-e^{-\lambda t},t\ge 0.$$

指数分布在可靠性领域中的应用

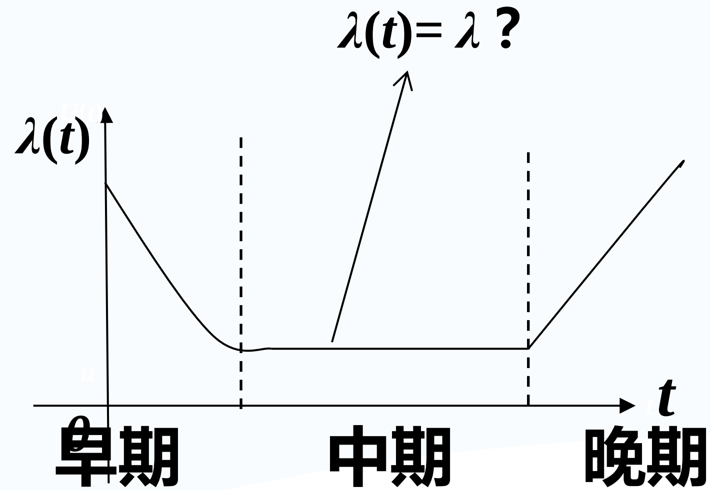
$$\lambda(t)=\lim_{\Delta t\to 0+}\frac{P\{T\le t+\Delta t\mid T>t\}}{\Delta t}$$

失效率函数的含义：由指数分布的参数引出可靠性中一类重要的函数，它决定每一时刻失效的强度。

失效率函数的应用：

从早期失效期、偶然失效期和磨损失效期三个阶段函数的变化，

给出在实际中产品性能对应的变化情况. 以工厂采取的筛选、维修、更换等措施对应的失效率函数的不同阶段来说明.



典型的失效率函数图(浴盆形)

指数分布在可靠性领域中的应用

$$\lambda(t)=\lim_{\Delta t\to 0+}\frac{P\{T\le t+\Delta t\,|\,T>t\}}{\Delta t}$$

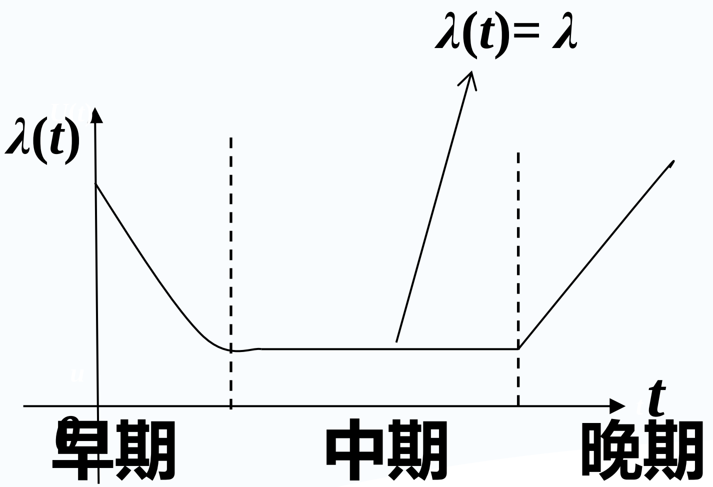
失效率函数的含义: 由指数分布的参数引出可靠性中一类重要的函数, 它决定每一时刻失效的强度.

失效率函数与指数分布的关系:

$$F(x)=1-C\cdot e^{-\int_0^t \lambda(u)du} \qquad (3.1)$$

$$F(0)=0$$

指数分布描述了“无老化”时的寿命分布,即处于稳定工作状态的寿命分布. 然而只是一种近似.



典型的失效率函数图(浴盆形)

高斯(Carl Friedrich Gauss, 1777—1855年)是德国数学家、物理学家和天文学家. 他和牛顿、阿基米德，被誉为有史以来的三大数学家.

高斯年仅三岁，就学会了算术，八岁因发现等差数列求和公式而深得老师和同学的钦佩. 1795年进入格丁根大学学习, 大学二年级时得出正十七边形的尺规作图法,并给出了可用尺规作图的正多边形的条件.解决了两千年来悬而未决的难题,1799年以代数基本定理的四个漂亮证明获博士学位.

高斯的数学研究几乎遍及所有领域. 他还把数学应用于天文学、大地测量学和磁学的研究, 发明了最小二乘法原理(第九章). 高理的数论研究 总结 在《算术研究》(1801)中, 这本书奠定了近代数论的基础, 它不仅是数论方面的划时代之作, 也是数学史上不可多得的经典着作之一.

正态分布的名称为“正态”(normal，意为通常的)，正态分布也称为“高斯分布”，正态分布的密度曲线也称为“钟形曲线”.正态分布是最重要的一种概率分布. 正态分布是一个在数学、物理及工程等领域都非常重要的概率分布，在统计学的许多方面有着重大的影响力.

正态分布概念是由德国的数学家和天文学家De Moivre在求二项分布的渐近公式中于1733年首次提出的(第五章中心极限定理涉及), C.F.高斯在研究测量误差(源于误差分析)时从另一个角度导出了它, P.S.拉普拉斯和高斯研究了它的性质. 由于德国数学家Gauss率先将其应用于天文学家研究，故正态分布又叫高斯分布.

高斯在1809年解决了误差分布问题, 给出误差服从正态分布. 正态分布作为一种概率模型, 在19世纪极为流行, 一些学者甚至把19世纪的统计学称为正态分布统治的年代.

高斯这项工作对后世的影响极大, 现今德国10马克的印有高斯头像的钞票上还印有正态分布的密度曲线.

生产与科学实验中很多随机变量的概率分布都可以近似地用正态分布来描述.例如, 在生产条件不变的情况下,产品的强力、抗压强度、口径、长度等指标；同一种生物体的身长、体重等指标；同一种种子的重量；测量同一物体的误差；弹着点沿某一方向的偏差；某个地区的年降水量；以及理想气体分子的速度分量等等.

一般来说，如果一个量是由许多微小(每个因素所起作用不大)的独立随机因素影响的结果，那么就可以认为这个量具有正态分布（见中心极限定理). 从理论上看,正态分布具有很多良好的性质, 许多概率分布可以用它来近似;还有一些常用的概率分布是由它直接导出的, 例如对数正态分布、*t*分布、*F*分布等.

教育统计学 统计规律表明,学生的智力水平,包括学习能力,实际动手能力等呈正态分布. 因而正常的考试成绩分布应基本服从正态分布.

例6：上海手表厂曾对其生产的某个零件的重量收集了大量资料，对测量得的3805个数据，按不同重量加以分组，并记录了不同范围内零件的个数(频数)，计算了它们的频数，结果如下表所示，它们与μ=56.94，σ=8.2的正态分布符合得相当好.

区间$[x_i,x_{i+1})$	频数 n_i	频率 $f_i=\frac{n_i}{n}$	$\Phi\left(\frac{x_{i+1}-\mu}{\sigma}\right)-\Phi\left(\frac{x_i-\mu}{\sigma}\right)$
$(-\infty,41.5)$	125	0.032 85	0.030 05
[41.5,43.5)	72	0.018 92	0.021 50
[43.5,45.5)	124	0.032 59	0.029 21
[45.5,47.5)	145	0.038 11	0.044 31
[47.5,49.5)	193	0.050 72	0.056 30
[49.5,50.5)	137	0.036 0	0.033 4
[50.5,51.5)	131	0.034 4	0.039 8
[51.5,52.5)	154	0.040 5	0.040 0
[52.5,53.5)	156	0.041 0	0.042 6
[53.5,54.5)	174	0.045 7	0.045 7
[54.5,55.5)	186	0.048 9	0.047 2
[55.5,56.5)	191	0.050 2	0.048 4
[56.5,57.5)	206	0.054 1	0.048 6
[57.5,58.5)	193	0.050 7	0.048 2
[58.5,59.5)	185	0.048 6	0.047 2
[59.5,60.5)	153	0.040 2	0.045 4
[60.5,61.5)	176	0.046 3	0.043 0
[61.5,62.5)	147	0.038 6	0.040 2
[62.5,63.5)	144	0.037 8	0.037 0
[63.5,64.5)	140	0.036 8	0.033 1
[64.5,65.5)	109	0.028 6	0.029 9
[65.5,66.5)	111	0.029 2	0.028 2
[66.5,67.5)	93	0.024 44	0.022 47
[67.5,69.5)	127	0.033 38	0.035 52
[69.5,71.5)	81	0.021 29	0.025 47
$[71.5,\infty)$	152	0.039 95	0.037 54
总和	$n=3\ 805$		

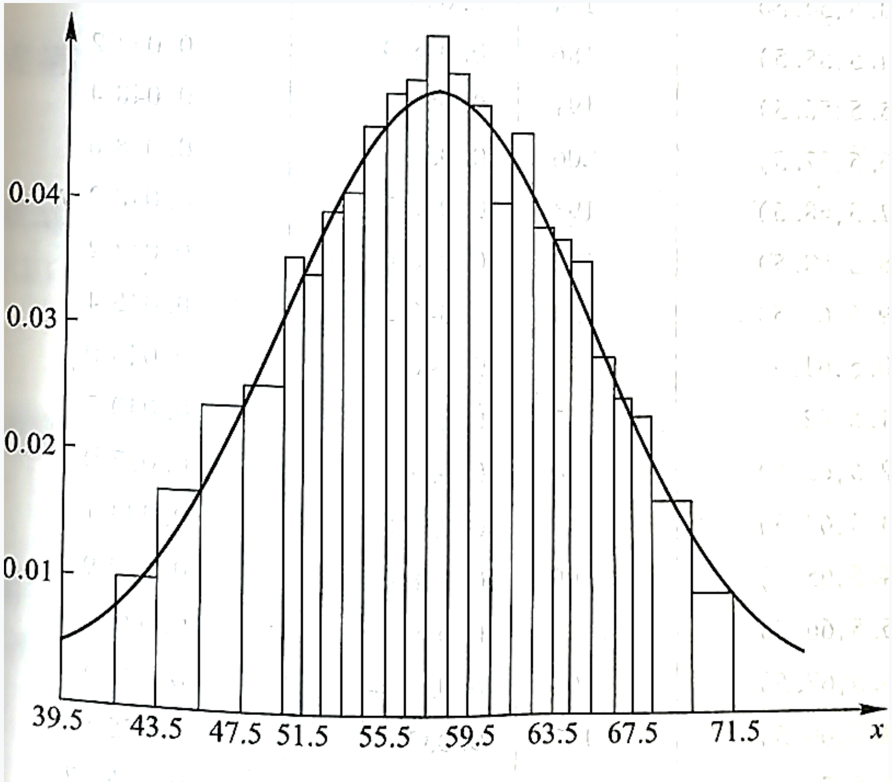


图 3.1.7 实测数据频率直方图与正态密度曲线

三：正态分布 (Gaussian分布)

设随机变量X 的概率密度函数为

$$\varphi(x;\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\quad x\in R$$

其中$\mu,\sigma\ (\sigma>0)$是常数，则称随机变量X 服从参数为 μ，σ^2 的正态分布(或高斯分布)，记为$X\sim N(\mu,\sigma^2)$

特别地, 当 $\mu=0$, $\sigma=1$时, 其概率密度函数为

$$\varphi(x)=\varphi(x;0,1)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\quad x\in R$$

则称随机变量X 服从标准正态分布, 即$X\sim N(0,1)$.

例7：设 $\varphi(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \quad x\in R$

则 $\varphi(x)$ 是概率密度函数.

证明：（1） $\varphi(x)>0, \ x\in R$ 显然成立.

(2) $\int_{-\infty}^{+\infty}\varphi(x)dx=1$

令 $I=\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}dx$

$$I^2=\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}dx\int_{-\infty}^{+\infty}e^{-\frac{y^2}{2}}dy$$

$$I^2=\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}dx\int_{-\infty}^{+\infty}e^{-\frac{y^2}{2}}dy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{x^2+y^2}{2}}dxdy$$

$$\sqrt{\pi}=\int_{-\infty}^{+\infty}e^{-y^2}dy$$

$$\Uparrow\ y=\frac{x}{\sqrt{2}}$$

$$\sqrt{2\pi}=\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}dx$$

$$\underline{\underline{\begin{matrix}x=r\cos\theta\\ y=r\sin\theta\end{matrix}}}\int_0^{2\pi}d\theta\int_0^{+\infty}e^{-\frac{r^2}{2}}rdr$$

$$=2\pi\left[-e^{-\frac{r^2}{2}}\right]_0^{+\infty}=2\pi$$

从而$I=\sqrt{2\pi}$ (概率积分)

所以 $\int_{-\infty}^{+\infty}\varphi(x)\,dx=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$

$$=\frac{1}{\sqrt{2\pi}}\sqrt{2\pi}=1$$

注: 采用耦合方法把一维问题转化为二维问题.

1　正态分布概率密度曲线的特征

(1) $\int_{-\infty}^{+\infty} \varphi(x;\mu,\sigma^2)\,dx = 1$

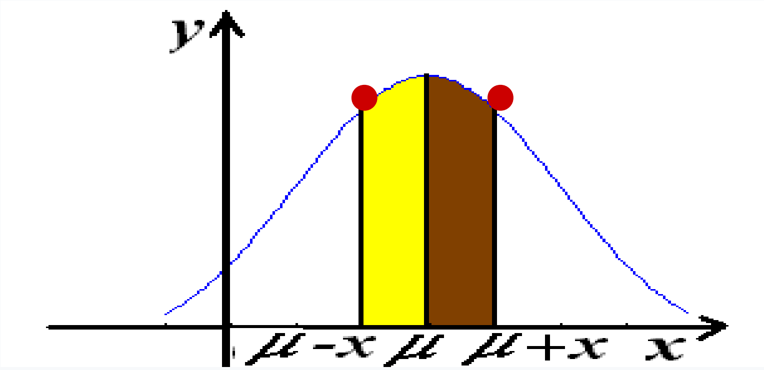
即概率曲线下总面积为1.

(2) 曲线关于直线 $x=\mu$ 对称，即对任意实数 x 有

$$\varphi(\mu - x;\mu,\sigma^2) = \varphi(\mu + x;\mu,\sigma^2)$$

曲线下直线 $x=\mu$ 两侧的面积各为1/2，

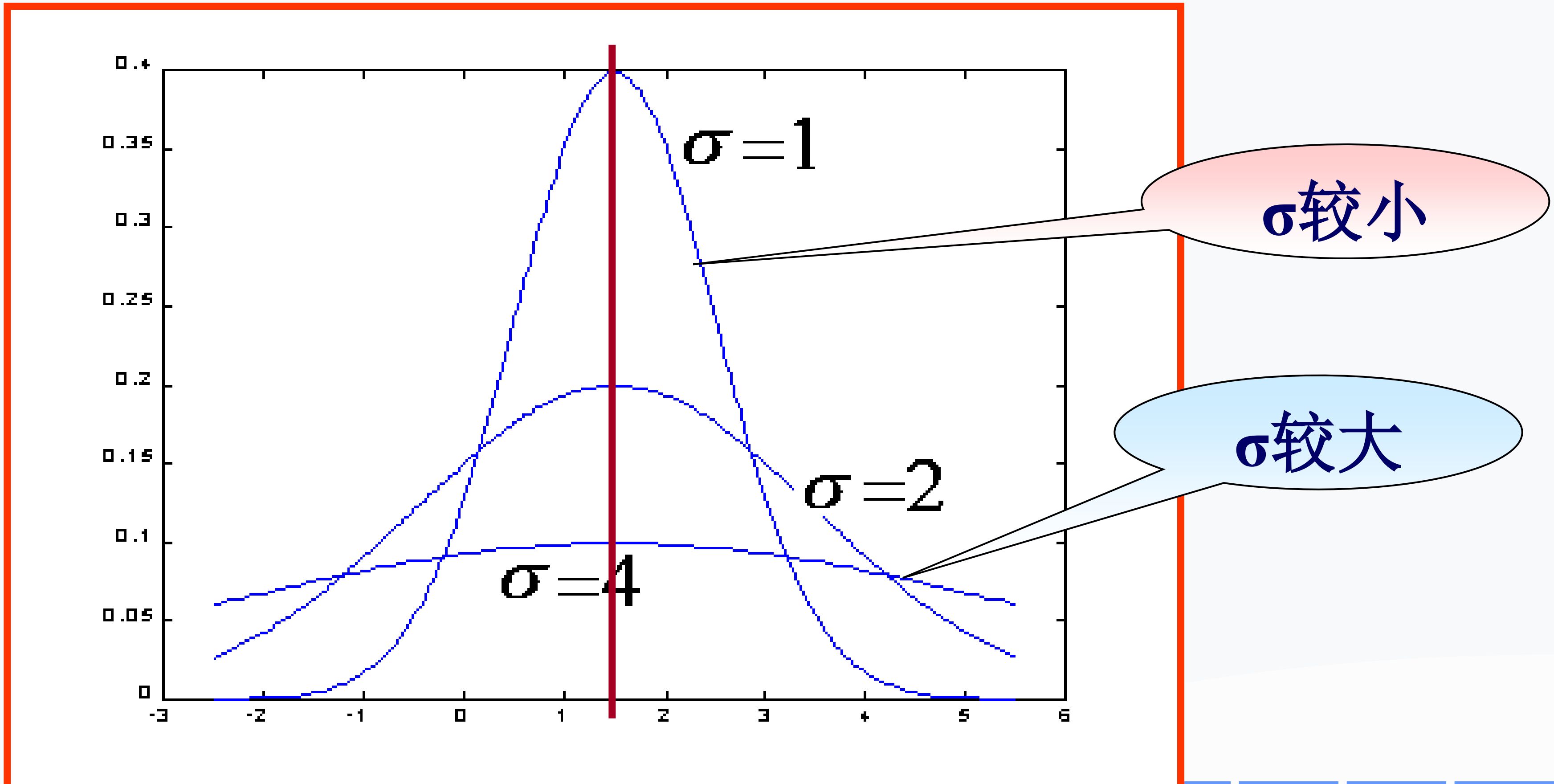
并且 $P\{\mu - x < X \le \mu\} = P\{\mu < X \le \mu + x)$

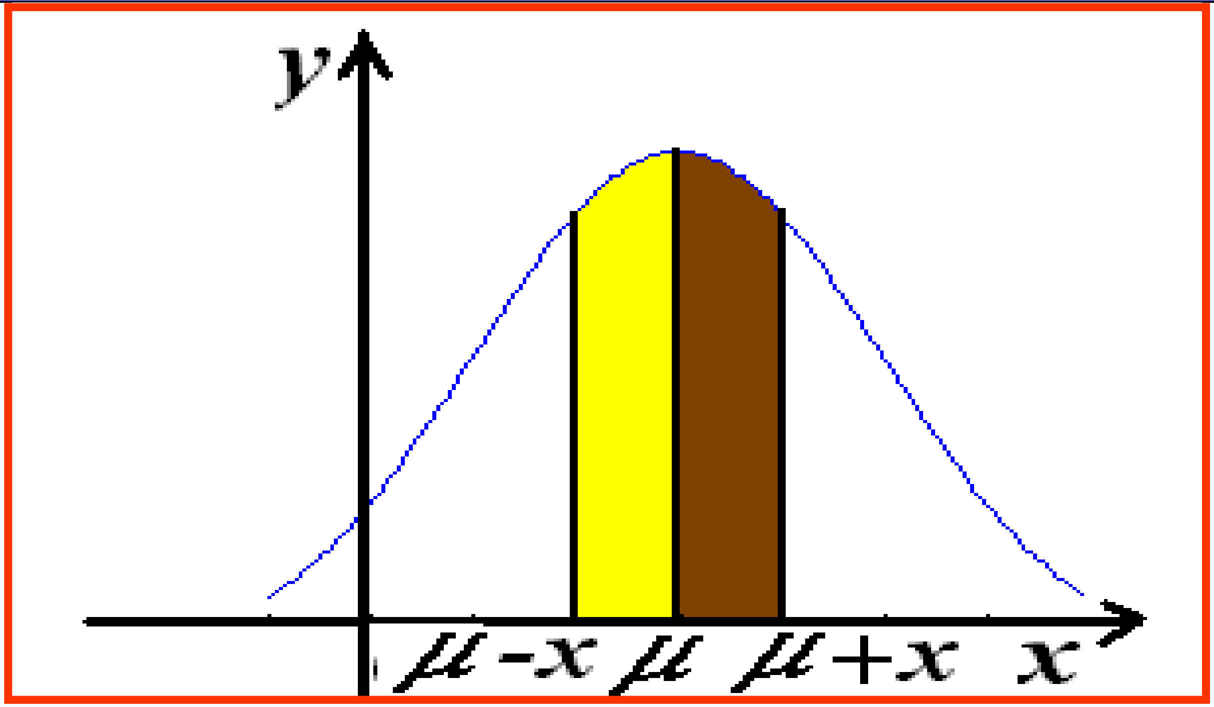


$$\varphi(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(3) 曲线 $x=\mu$ 处取得最大值 $\dfrac{1}{\sigma\sqrt{2\pi}}$， $\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

固定 μ，σ^2 越大，曲线越趋于平坦.





1. 若$X\sim N$（$2, \sigma^2$），且$P\{2<X<4\}=0.3$,

则$P\{0<X\}=$________________.

2. 若$X\sim N(\mu,\sigma^2)$, 且

$P\{3<X\}=P\{X<1\}$, 则$\mu=$?

2 正态分布概率的计算

若随机变量$X \sim N(\mu, \sigma^2)$，其分布函数为

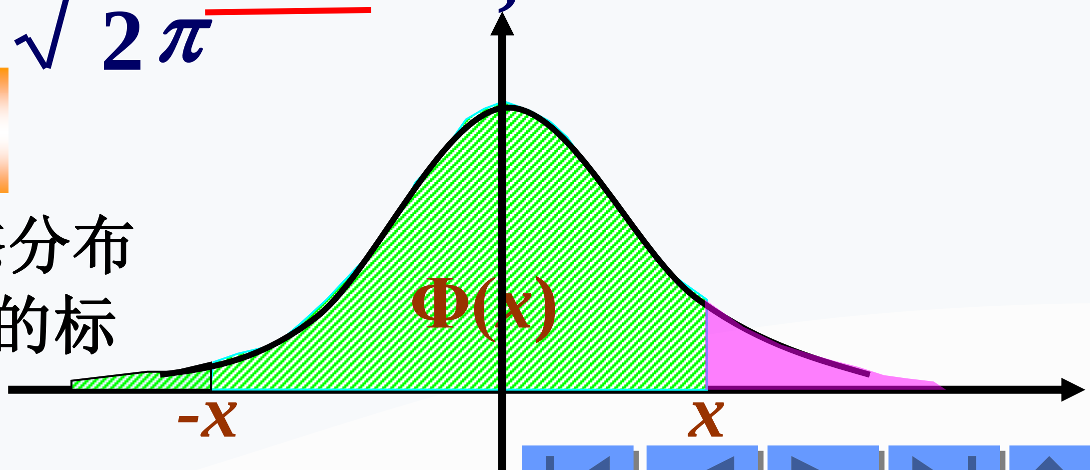
$$\Phi(x;\mu,\sigma^2)=\int_{-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt,\ x\in R$$

若随机变量X标准正态分布, 其分布函数为

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt,\ x\in R$$

$$\Phi(-x)=1-\Phi(x)$$

人们编制了《标准正态分布表》(见附表2), 给出了$x\geq 0$ 的标准正态分布函数值.



（1）若随机变量$X\sim N(0，1)$, 则

$$P\{a<X\le b\}=\Phi(b)-\Phi(a)$$

（2）若随机变量$X\sim N(\mu，\sigma^2)$，则

$$P\{x_1<X\le x_2\}=\Phi\left(\frac{x_2-\mu}{\sigma}\right)-\Phi\left(\frac{x_1-\mu}{\sigma}\right)$$

证明： 目的：$\Phi(x;\mu,\sigma^2)=\Phi\left(\frac{x-\mu}{\sigma}\right)$

$$\Phi(x;\mu,\sigma^2)=\int_{-\infty}^{x}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt$$

$$\overset{y=\frac{t-\mu}{\sigma}}{\underset{dt=\sigma dy}{=\!=\!=\!=}}\int_{-\infty}^{\frac{x-\mu}{\sigma}}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy=\Phi\left(\frac{x-\mu}{\sigma}\right)$$

正态分布事件概率

例8：已知随机变量度$X \sim N(\mu, \sigma^2)$，证明

$$P\{|X-\mu|<x\} = 2\Phi(\frac{x}{\sigma})-1$$

证明:

$$P\{|X-\mu|<x\} = P\{\mu - x < X < \mu + x\}$$

$$= \Phi(\frac{\mu + x - \mu}{\sigma}) - \Phi(\frac{\mu - x - \mu}{\sigma})$$

$$= \Phi(\frac{x}{\sigma}) - \Phi(\frac{-x}{\sigma})$$

$$= \Phi(\frac{x}{\sigma}) - [1 - \Phi(\frac{x}{\sigma})] = 2\Phi(\frac{x}{\sigma}) - 1$$

例8：已知随机变量度$X \sim N(\mu, \sigma^2)$，证明

$$P\{|X-\mu|<x\} = 2\Phi(\frac{x}{\sigma}) - 1$$

特别地, 有

$$P\{|X-\mu|<\sigma\} = 2\Phi(1) - 1 = 2\Phi(1.00) - 1 = 0.6826$$

$$P\{|X-\mu|<2\sigma\} = 2\Phi(2) - 1 = 2\Phi(2.00) - 1 = 0.9544$$

$$P\{|X-\mu|<3\sigma\} = 2\Phi(3) - 1 = 2\Phi(3.0) - 1 = 0.9974$$

这说明X以很大的概率密集在$x=\mu$的附近.

3σ原则

一般可以认为, X不在$(\mu-3\sigma,\mu+3\sigma)$外取值, 称为正态分布的“$3\sigma$”原则.

标准正态分布性质

(1) $\Phi(0)=\frac{1}{2}$

(2)若 $X\sim N(\mu,\sigma^2)$，则 $\frac{X-\mu}{\sigma}\sim N(0,1)$ (标准化处理)

$$P\{x_1<X\le x_2\}=P\{\frac{x_1-\mu}{\sigma}<\frac{X-\mu}{\sigma}\le\frac{x_2-\mu}{\sigma}\}$$

$$=\Phi(\frac{x_2-\mu}{\sigma})-\Phi(\frac{x_1-\mu}{\sigma})$$

(3) 当$X>3.9$时, 认为$\Phi(x)\approx 1$

(4)对一切$x\ge 0,\Phi(-x)=1-\Phi(x)$

注：常用的$\sqrt{2\pi}=\int_{-\infty}^{+\infty}e^{-\frac{x^2}{2}}dx$

$$\overset{y=\frac{x}{\sqrt{2}}}{\Rightarrow}\sqrt{\pi}=\int_{-\infty}^{+\infty}e^{-y^2}dy$$

$$\int_{-y}^{y}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\Phi(y)-\Phi(-y)$$

题

1.设连续型随机变量X的分布函数$F(x)=\begin{cases} A+Be^{-\frac{x^2}{2}}, & x\geq 0 \\ 0, & x<0 \end{cases}$

(1)求常数A,B;

2.设连续型随机变量X的分布函数为

$$F(x)=\begin{cases} 0, & x<0; \\ Ax^2, & 0\leq x\leq 1; \\ 1, & x\geq 1. \end{cases}$$

则A_____, X的密度函数为_____。

3. $X\sim N(1,4)$, $P\{X>c\}=2P\{X\leq c\}$,

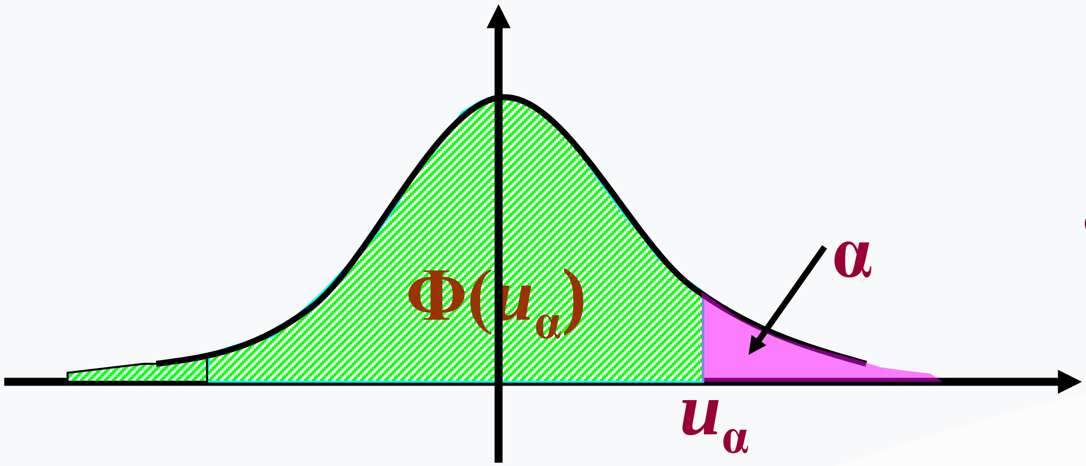
有$c=$______,其中$\Phi(0.43)=\dfrac{2}{3}\approx 0.6667.$

有时，我们需要求随机变量以给定概率落在某个区间上的分界点，称之为分位数.

分位数　　$X \sim N(0, 1)$，若实数u_α使

$$P\{X > u_\alpha\} = \alpha$$

则称u_α为标准正态分布的对应于α的上侧分位数.



$\Phi(u_\alpha) = 1-\alpha$

分位数

例9：设$X \sim N(10, 4)$，求α使 $P\{|X - 10| < \alpha\} = 0.9$

解:

$$P\{|X-10|<\alpha\} = P\{10-\alpha < X < 10+\alpha\}$$

$$= \Phi\left(\frac{10+\alpha-10}{2}\right) - \Phi\left(\frac{10-\alpha-10}{2}\right)$$

$$= \Phi\left(\frac{\alpha}{2}\right) - \left[1-\Phi\left(\frac{\alpha}{2}\right)\right] = 2\Phi\left(\frac{\alpha}{2}\right) - 1 = 0.9$$

$$\Rightarrow \Phi\left(\frac{\alpha}{2}\right) = 0.95$$

$$\Rightarrow \frac{\alpha}{2} = 1.645$$

$$\Rightarrow \alpha = 3.29$$

例10:用正态分布估计高考录取最低分. 某市有9万名高中毕业生参加高考, 结果有5.4万名被各类高校录取. 已知满分为600分, 540分以上者有2025人, 360分以下者有13500人, 试估计高考录取最低分.

分析:考生的高考成绩为随机变量X, 它一般会受先天遗传、后天努力、心理素质、考试期间身体状态、求学期间班级学风及有无请家教等诸多随机因素的影响, 而各因素的影响是有限的, 但正、负影响会相互抵消, 故X服从参数为μ, σ^2 的正态分布.

(1)通过高考结果的两个信息, 建立关于未知参数μ, σ^2 的两个方程;

(2)通过已公布的录取率, 求得录取的最低分.

解:取设学生高考成绩$X \sim N(\mu, \sigma^2)$, 则

$$P\{X \le 540\} = 1\text{-}P\{X > 540\} = 1 - \frac{2025}{90000} = 0.9775,$$

$$P\{X \le 540\} = \Phi\left(\frac{540-\mu}{\sigma}\right) = 0.9775,$$ 查表 $\frac{540-\mu}{\sigma}$=2.05(修正)

例10：用正态分布估计高考录取最低分. 某市有9万名高中毕业生参加高考, 结果有5. 4万名被各类高校录取. 已知满分为600分, 540分以上者有2025人, 360分以下者有13500人, 试估计高考录取最低分.

$$P\{X<360\}=\frac{13500}{90000}=0.15=\Phi(\frac{360-\mu}{\sigma}),\quad \Phi(\frac{\mu\text{-}360}{\sigma})=1-0.15=0.85$$

查正态分布表, 得 $\dfrac{540-\mu}{\sigma}$ =2.05(修正), $\dfrac{\mu-360}{\sigma}=1.04$,

解上述方程组: $\mu\approx 421,\ \sigma\approx 58.$

$$\Phi(-x)=1-\Phi(x)$$

已知录取率 $\dfrac{54000}{90000}$ =0.6, 设被录取者最低分为a，则

$$P\{X\geq a\}=0.6=1\text{-}P\{X<a\}=1-\Phi(\frac{a-421}{58})=\Phi(\frac{421-a}{58}),$$

查正态分布表, 即得 $\dfrac{421-a}{58}$ =0.255(修正) $\Rightarrow a\approx 406$

VaR(Value at Risk) 按字面解释就是"在险价值"，其含义指：在市场正常波动下,某一金融资产或证券组合的最大可能损失。

更为确切的是指，在一定概率水平(置信度)下，某一金融资产或证券组合价值在未来特定时期内的最大可能损失量。

用公式表示为：

P(ΔPΔt≤VaR)=a

P——资产价值损失小于可能损失上限的概率，即英文的Probability。

ΔP——某一金融资产在一定持有期Δt的价值损失额。

VaR——给定置信水平a下的在险价值，即可能的损失上限。

a——给定的置信水平

"Given some confidence level $\alpha \in (0,1)$ the VaR of the portfolio at the confidence level α is given by the smallest number l such that the probability that the loss L exceeds l is not larger than (1 − α)"[2]

$$VaR_\alpha = \inf\{l \in \Re : P(L > l) \le 1 - \alpha\} = \inf\{l \in \Re : F_L(l) \ge \alpha\}$$

The left equality is a definition of VaR. The right equality assumes an underlying probability distribution, which makes it true only for parametric VaR.

VaR从统计的意义上讲，本身是个数字，是指面临“正常”的市场波动时“处于风险状态的价值”。
能力来确定。

例如，某一投资公司持有的证券组合在未来24小时内，置信度为95%，在证券市场正常波动的情况下，VaR值为520万元，其含义是指，该公司的证券组合在一天内(24小时)，由于市场价格变化而带来的最大损失超过520万元的概率为5%，平均20个交易日才可能出现一次这种情况。

或者说有95%的把握判断该投资公司在下一个交易日内的损失在520万元以内。

5%的几率反映了金融资产管理者的风险厌恶程度，可根据不同的投资者对风险的偏好程度和承受能力来确定。

"Given some confidence level $\alpha \in (0,1)$ the VaR of the portfolio at the confidence level α is given by the smallest number l such that the probability that the loss L exceeds l is not larger than $(1-\alpha)$"[2]

$$VaR_\alpha = \inf\{l \in \Re : P(L > l) \le 1-\alpha\} = \inf\{l \in \Re : F_L(l) \ge \alpha\}$$

传统的ALM(Asset-Liability Management,资产负债管理)过于依赖报表分析，缺乏时效性；利用方差及β系数来衡量风险太过于抽象，不直观，而且反映的只是市场（或资产）的波动幅度；

而CAPM(资本资产定价模型)又无法揉合金融衍生品种。在上述传统的几种方法都无法准确定义和度量金融风险时，G30集团在研究衍生品种的基础上, 于1993年发表了题为《衍生产品的实践和规则》的报告，提出了度量市场风险的VaR(Value at Risk：风险价值)方法已成为目前金融界测量市场风险的主流方法。

稍后由J.P.Morgan推出的用于计算 VaR的Risk Metrics风险控制模型更是被众多金融机构广泛采用。目前国外一些大型金融机构已将其所持资产的VaR风险值作为其定期公布的会计报表的一项重要内容加以列示。