目录

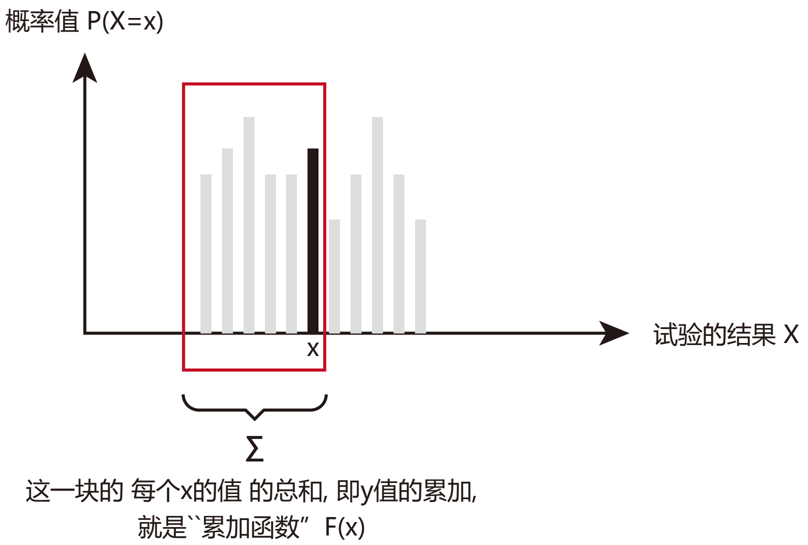
文件名

第一部分　概率函数：PMF (离散型数据的, 称为"概率质量函数 Probability mass function") & PDF (连续性数据的, 称为"概率密度函数 Probability Density Function")

第二部分　累积函数 $F(x) = P\{X \leq x\} = \sum_{x_k \leq x} p_k$ ：Cumulative Distribution Function (CDF) ← 是对"概率函数"值的累加结果

对于随机变量, 我们通常关心的, 并不是它取某个值的概率(即我们并不关心它的分布律), 而是更关心它落在某个区间内的概率.
比如, 对某考试, 我们更关心的是"不及格的总人数", 和比如 "分数≥80分的总人数".

累积函数 Cumulative Distribution Function (CDF) ← 是对"概率函数"值的累加结果 . 即对"概率密度函数"的积分.

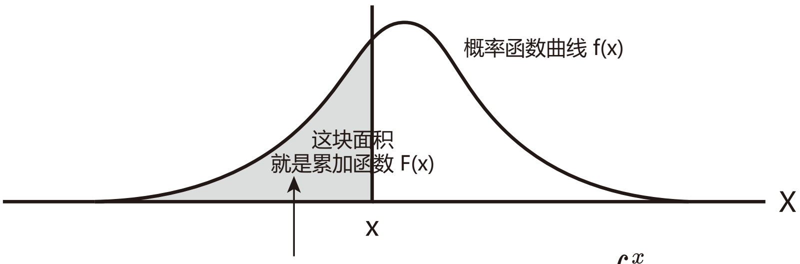


在这些个区间段所占的概率值, 就是用"累加函数"(又叫"分布函数")来表示的. 即:

$(\text{随机变量}X \leq \text{自变量}x) = \underbrace{F(x)}_{\text{累加函数}} \leftarrow$ 它表示随机变量X 落在 (-∞, x] 这段区间上的概率.

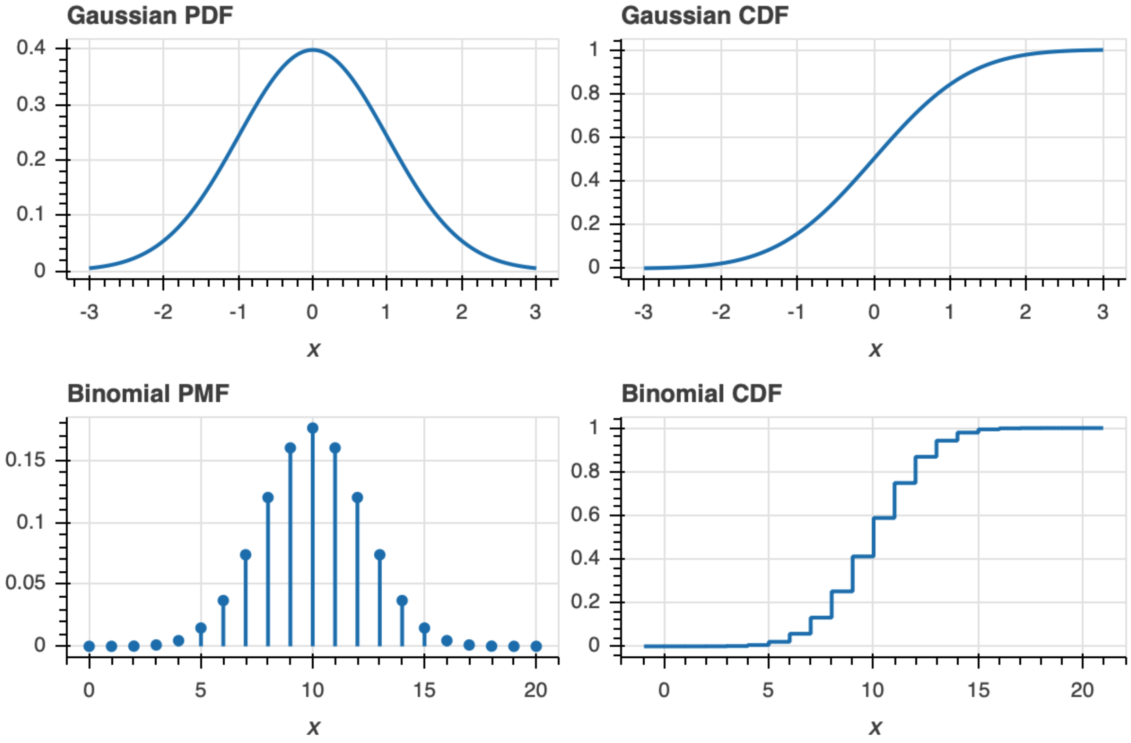
$$\boxed{\text{累加函数}F(x) = P\{X \leq x\} = \sum_{x_k \leq x} p_k}$$

累加函数 F(x) 就是 “X取≤x 的所有值x_k” 的概率之和.
P(X≤x) 即“X的取值不超过x” 的概率. 这里P后面写()或{}都行, 意思是一样的.



$$F(x) = P(X \leq x) = S_{\text{阴影部分面积}} = \int_{-\infty}^{x} f(t) d$$

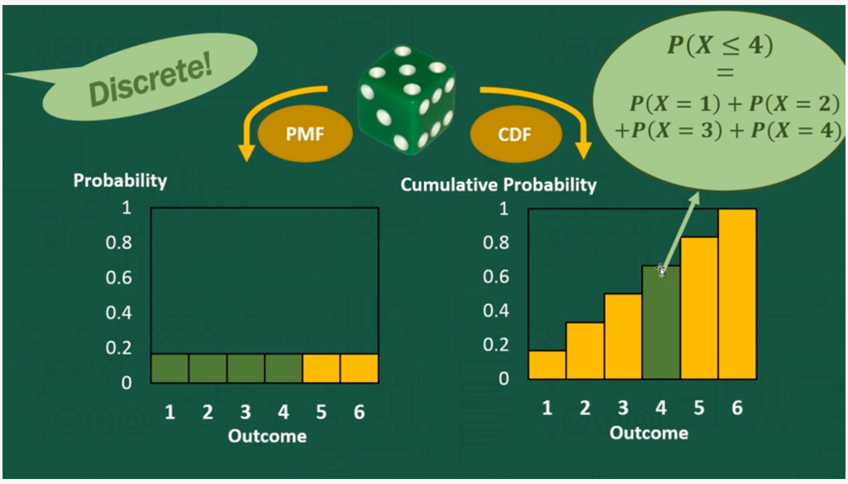
下图, 左边两张是“概率函数”, 右边两张就是“累加函数 CDF”.



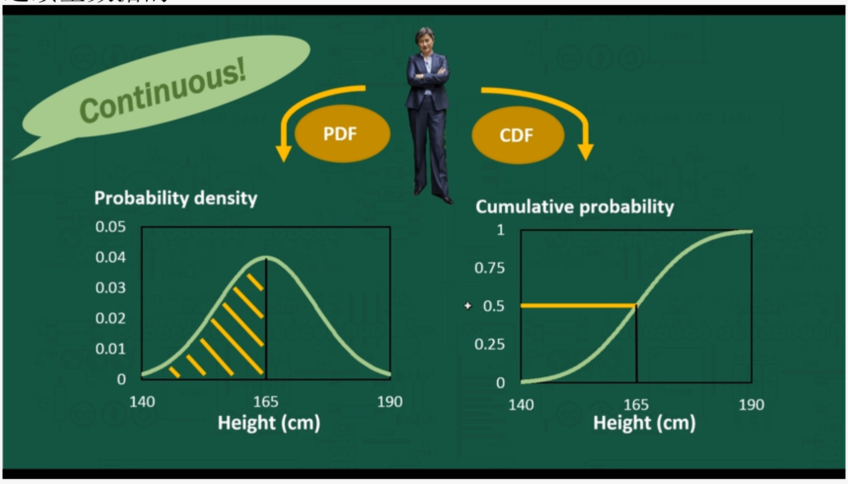
例

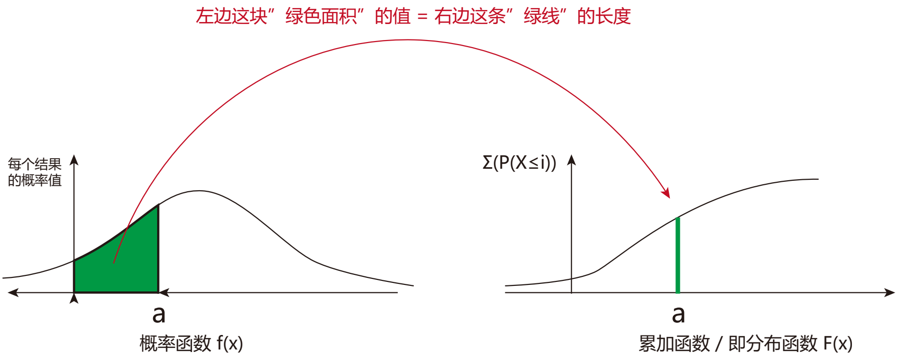
下面的图, 左边是“概率函数”, 右边是“累加函数”.

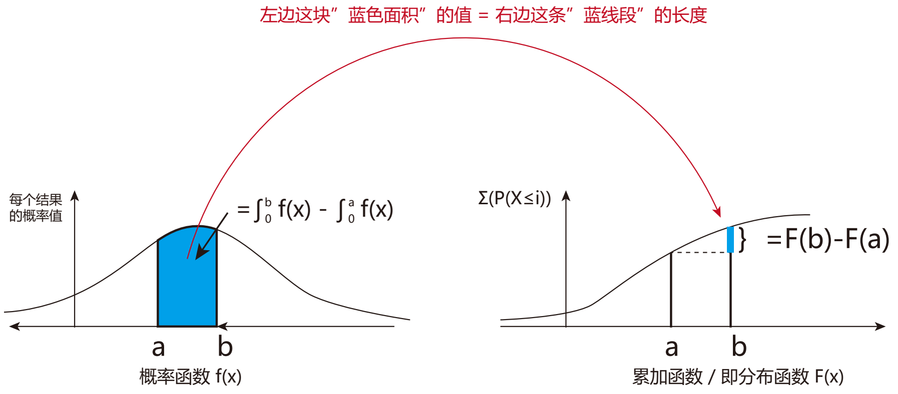
离散型数据的:

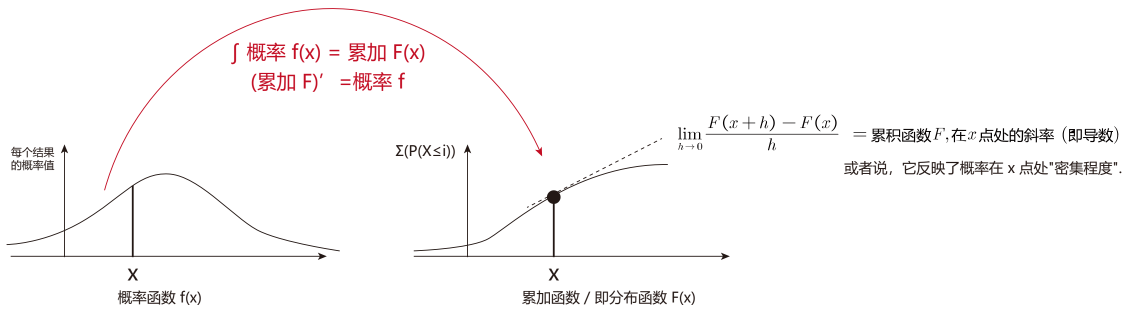

Discrete!
PMF
CDF
P(X ≤ 4)
=
P(X = 1) + P(X = 2)
+P(X = 3) + P(X = 4)
Probability
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6
Outcome
Cumulative Probability
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6
Outcome

连续型数据的:

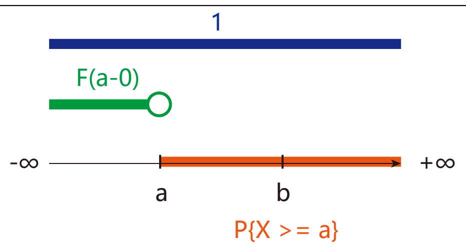
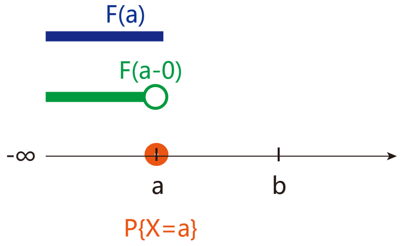
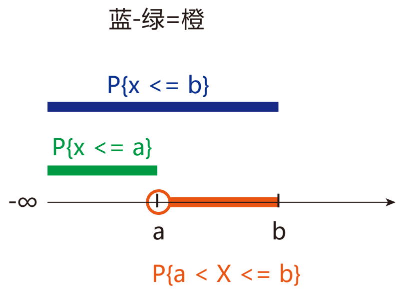
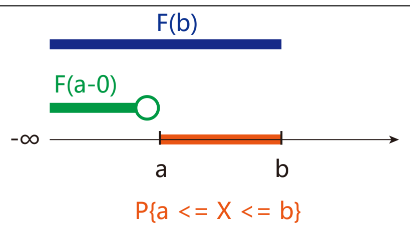

Continuous!
PDF
CDF
Probability density
0.05
0.04
0.03
0.02
0.01
0
140
165
190
Height (cm)
Cumulative probability
1
0.75
0.5
0.25
0
140
165
190
Height (cm)


左边这块”绿色面积”的值 = 右边这条”绿线”的长度
每个结果
的概率值
a
概率函数 f(x)
Σ(P(X≤i))
a
累加函数 / 即分布函数 F(x)


左边这块”蓝色面积”的值 = 右边这条”蓝线段”的长度
每个结果
的概率值
= ∫₀ᵇ f(x) - ∫₀ᵃ f(x)
a b
概率函数 f(x)
Σ(P(X≤i))
} =F(b)-F(a)
a b
累加函数 / 即分布函数 F(x)

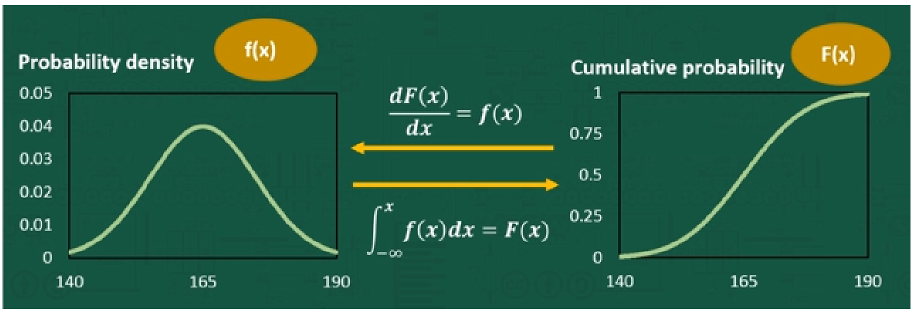


1 ★累加函数的计算公式

累加函数 $F(x) = PX \le x$ 的公式有	图中: 蓝-绿=橙
(1) $P\{X \le a\} = F(a)$	
(2) $P\{X < a\} = F(a-0)$ ← 其中的 F(a-0)：就是从左边逼近a, 不包括a点. 所以是“左极限”. 就是(-∞,a)这段区间的概率之和, 不包括a点上的概率.	
(3) $P\{X > a\} = 1 - P\{X \le a\} = 1 - F(a)$	
(4) $P\{X \ge a\} = 1 - F(a-0)$	
(5) $P\{X = a\} = F(a) - F(a-0)$	
(6) $P\{a < X \le b\} = P(X \le b) - P(X \le a)$	
(7) $P\{a \le X \le b\} = F(b) - F(a-0)$	

2 对“概率函数f(x)”求积分, 就得到“累加函数F(x)”. 反之, 对“累加函数F(x)”求导, 就得到“概率函数f(x)”. 即: ① $\int$ 概率函数$f(x)$ = 累加函数$F(x)$, ② (累加函数$F(x))'$ = 概率函数$f(x)$

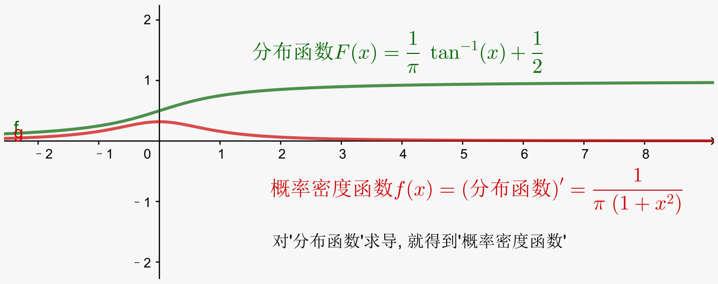
对“概率函数f(x)”求积分, 就得到“累加函数 F(x)”
对“累加函数 F(x)” 求导, 就得到“概率函数f(x)”.



例

$$\underbrace{f(x)}_{\text{概率函数(相当于导函数)}} = \frac{1}{\pi(1+x^2)}$$

$$\underbrace{F(x)}_{\text{累加函数}} = \int_{-\infty}^{x} \underbrace{f(t)}_{\text{概率函数}} dt = \int_{-\infty}^{x} \underbrace{\frac{1}{\pi(1+t^2)}}_{\text{导函数}} dt = \frac{1}{\pi}\left(\arctan|_{-\infty}^{x}\right) = \frac{1}{\pi}\arctan x + \frac{1}{2}$$



例

有累加函数 $F(x) = \begin{cases} 0 & (x<0) \\ Ax^2 & (0 \le x < 1) \\ 1 & (x \ge 1) \end{cases}$

→ 求常数A : $\underbrace{\lim_{x\to 1^-} Ax^2}_{\text{从}x=1\text{的左侧,逼近}x=1} = A(1)^2 = A = \underbrace{F(1)=1}_{\text{因为当}x\ge 1\text{时,}F(x)=1}$

→ 求概率函数f(x) : $f(x) = F'(x) = \begin{cases} 0' = 0 & (x<0) \\ (Ax^2)' = (x^2)' = 2x & (0 \le x < 1) \\ 1' = 0 & (x \ge 1) \end{cases}$

→ 求：

$$P\{0.3 < X < 0.7\} = \underbrace{F(0.7)}_{=\int_{-\infty}^{0.7} f(t)dt} - \underbrace{F(0.3)}_{=\int_{-\infty}^{0.3} f(t)dt}$$

因为在$0 \le x < 1$的区间上,$F(x) = Ax^2$,而其中的A我们上面已经算出$=1$,

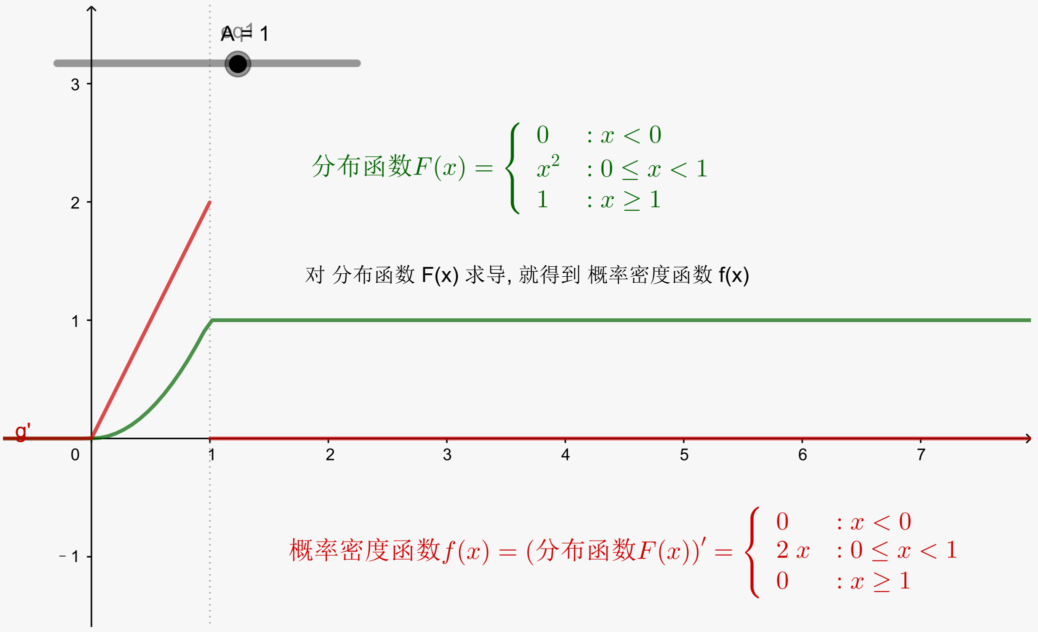
所以$F(x) = Ax^2 = (1)x^2 = x^2$

所以: $F(0.7) = (0.7)^2 = 0.49$

$F(0.3) = (0.3)^2 = 0.09$

因此: $P\{0.3 < X < 0.7\} = F(0.7) - F(0.3) = 0.49 - 0.09 = 0.4$

事实上,本例的 $P\{0.3 < X < 0.7\} = \int_{0.3}^{0.7} \underbrace{(2x)}_{\text{即概率函数}f(x)} dx$

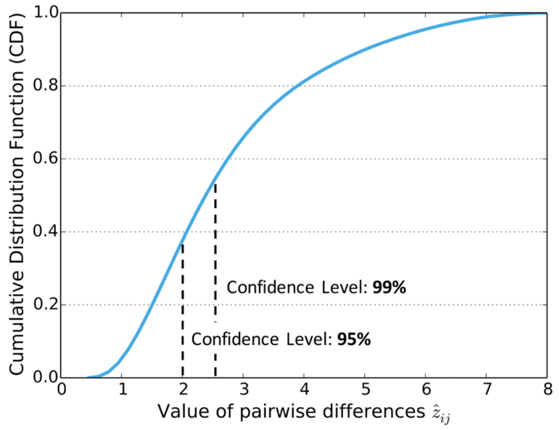


---

3 性质

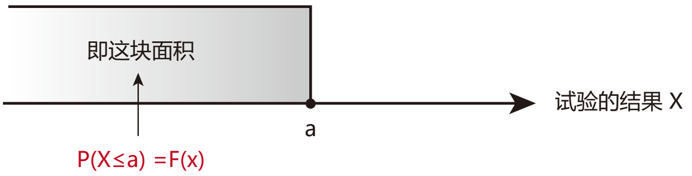
3.1 性质1: 有界性. $F(x) = P\{随机变量X \le 随机变量的取值x\}$, $x \in (-\infty, +\infty)$. 即 0 $F(x) \le 1$

累加函数(CDF) F(x), 就是一个普通的实函数. 其定义域是 $x \in (-\infty, +\infty)$. 值域是 $y \in [0, 1]$.

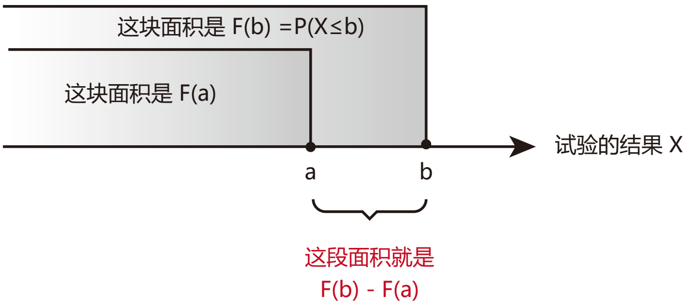


累加函数：

$$\boxed{F(x) = P(X \le a) = \int_{lower\lim it}^{a} f(x)\,dx}$$



$$\boxed{P(a \le X \le b) = P(a < X < b) = \int_a^b f(x)\,dx = F(b) - F(a)}$$



$P(x_1 < X \le x_2)$ ¸ 对于随机变量X在$(x_1, x_2]$这段区间上的概率,它的值

$= F(x_2) - F(x_1)$

$= P\{X \le x_2\} - P\{X \le x_1\}$

对于“连续型随机变量”，有没有两端的端点，无所谓，不影响概率值(因为它在任何一个“确定点”的概率都是0嘛)．即:

$$\begin{aligned}
&P\{a \le X \le b\} \\
&= P\{a < X \le b\} \quad \leftarrow \text{即，两端是否有“等于号”，无所谓.} \\
&= P\{a \le X < b\} \\
&= P\{a < X < b\}
\end{aligned}$$

同样:

$P\{X < a\} = P\{X \le a\} \leftarrow$ 有没有“等于号”无所谓

$P\{X > a\} = P\{X \ge a\}$

3.2 性质2: 单调不减性. 即对于任意的 $x_1 < x_2$, 有: $F(x_1) \leq F(x_2)$

F(x)是关于x的“不减函数”, 类似于“单调递增”的概念. “不减”的意思就是, 该函数的y值不会下降, 只会“增长”或“平移向前”.

比如, “分数小于等于70分的人” 其概率一定是小于等于 “分数小于80分的人”. 即 $F(70) \leq F(80)$.

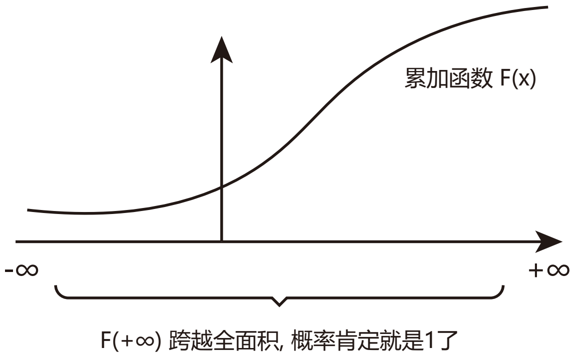
3.3 性质3: 规范性. F(-∞)=0 , F(+∞)=1

$\boxed{\underbrace{F(-\infty)}_{=P(X\leq-\infty)} = \lim_{x\to-\infty} F(x) = P(X < -\infty) = P(\Phi) = 0}$ ← 称之为 “不可能事件”.

如果随机变量X的取值, 比 -∞还小, 那其概率, 就只能是0了.

$\boxed{\underbrace{F(+\infty)}_{=P(X\leq+\infty)} = \lim_{x\to+\infty} F(x) = P(X < +\infty) = P(\Omega) = 1}$ ← 称之为 “必然事件”.

如果随机变量X的取值, 在 +∞以下, 那其概率, 肯定就是100%了, 就是1.



3.4 性质4 : 右连续性. $\lim_{x\to x_0^+} F(x) = F(x_0)$

这个等式的意思就是说: 累加函数在x_0点的右极限, 就等于累加函数在该点处的函数值.

右连续	所谓“右连续”, 就是“函数从x在某点的右侧, 逼近该点”的极限值, 就等于“该点处的y值”, 即: $\lim_{x\to a^+} F(x) = F(a)$.
左连续	同理, “左连续”就是: $\lim_{x\to a^-} F(x) = F(a)$.
连续	同时满足“左连续”和“右连续”的函数, 就称为是“连续”的. 即 $\lim_{x\to a} F(x) = F(a)$.

满足上面4条性质的, 就一定是“累加函数”. 反之, “累加函数”也一定有这4条性质.

第三部分　随机变量函数的分布

意思就是说, 假如我们已经知道某个X 是某种类型的分布了, 比如 X 它是几何分布的, 二项分布的等. 则进一步, 而我们还想知道, 用这个X 来构造出的其他函数, 会是什么类型的分布呢? 比如, Y=3X-5, 这个Y是由X构造出来的, 那么这个Y, 也是和X相同类型的分布吗? 还是说, Y是其他类型的分布?

$\rightarrow$ 随机变量X, 它取x时, 其“累加函数”是: $\boxed{F_X(x) = P\{X \leq x\}}$

$\rightarrow$ 由随机变量X, 构造出的一个新 Y (比如 Y=“多少倍的X, 再加上某个数”之类), 这个Y 的“累加函数”, 是: $\boxed{F_Y(x) = P\{Y \leq x\}}$ $\leftarrow$ 等号左边的 $F_Y(x)$ 意思是: Y是从X构造出来的. 累加函数(用F表示), 所以 $F_Y(x)$ 就是指 “由X构造出来的新的Y”的累加函数.