

Ch04: 连续型随机变量

Typical Continuous Variables and Examples

October 17, 2023

常用连续型随机变量 – 均匀分布

定义 0.31 若连续随机变量 X 的密度函数为

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b] \\ 0, & \text{其它} \end{cases}$$

称 X 服从区间 $[a, b]$ 上的均匀分布, 记为 $X \sim U(a, b)$.

若随机变量 $X \sim U(a, b)$, 则 X 落入内任一子区间 $[x, x + \Delta]$ 的概率

$$P(x \leq X \leq x + \Delta) = \int_x^{x+\Delta} \frac{1}{b-a} dt = \frac{\Delta}{b-a}$$

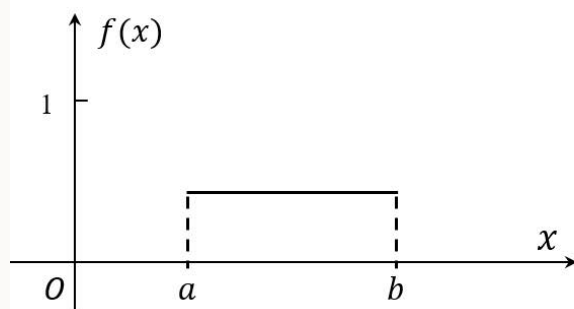
几何解释: 若 X 落入 $[a, b]$ 内任一子区间的概率与该区间的长度成正比, 与位置无关.

均匀分布的分布函数

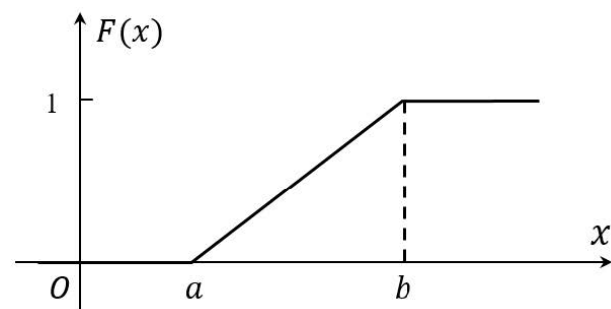
若随机变量 $X \sim U(a, b)$ 的分布函数为

$$F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a < x < b \\ 1, & x \geq b \end{cases}$$

随机变量 $X \sim U(a, b)$ 的密度函数和分布函数的示意图为:



密度函数



分布函数

均匀分布的期望与方差

若随机变量 $X \sim U(a, b)$, 则

$$\mathbb{E}(X) = \frac{a+b}{2} \quad \mathbb{VAR}(X) = \frac{(b-a)^2}{12}$$

均匀分布：例 0.68

例 **0.68** 已知随机变量 $X \sim U(a, b)$, 则 $a < c < d < b$, 求 $P(X \leq c \mid X \leq d)$.

解答：例 0.68

题目：已知随机变量 $X \sim U(a, b)$, 则 $a < c < d < b$, 求 $P(X \leq c | X \leq d)$.

解答：

- 根据条件概率的定义及均匀分布的密度函数可知,

$$P(X \leq c | X \leq d) = \frac{P(\{X \leq d\} \cap \{X \leq c\})}{P(X \leq d)} = \frac{P(X \leq c)}{P(X \leq d)} = \frac{c - a}{d - a}$$

即在 $X \leq d$ 的条件下, 随机变量 $X \sim U(a, d)$.

均匀分布：例 0.69

例 0.69 设随机变量 $\xi \sim U(-3, 6)$, 试求方程 $4x^2 + 4\xi x + \xi + 2 = 0$ 有实根的概率.

解答：例 0.69

题目：设随机变量 $\xi \sim U(-3, 6)$ ，试求方程 $4x^2 + 4\xi x + \xi + 2 = 0$ 有实根的概率。

解答：

- 根据均匀分布的密度函数和一元二次方程有实数根的条件可知，随机变量 ξ 的密度函数为

$$f(x) = \begin{cases} \frac{1}{9}, & x \in [-3, 6] \\ 0, & \text{其它} \end{cases}$$

设事件 A 表示方程有实根，于是有，

$$\begin{aligned} P(A) &= P((4\xi)^2 - 4 \times 4 \times (\xi + 2) \geq 0) = P((\xi + 1)(\xi - 2) \geq 0) \\ &= P(\{\xi \geq -1\} \cap \{\xi \geq 2\} \geq 0) + P(\{\xi \leq -1\} \cap \{\xi \leq 2\} \geq 0) \\ &= P(\xi \leq -1) + P(\xi \geq 2) = \int_{-3}^{-1} \frac{1}{9} dt + \int_2^6 \frac{1}{9} dt = \frac{2}{3} \end{aligned}$$

常用连续型随机变量 – 指数分布

指数分布常用于电话的通话时间和银行的服务等待时间,也可以用于描述动物和电子元件的寿命,在可靠性理论和排队论中具有广泛的应用.

定义 0.32 给定常数 $\lambda > 0$, 若连续随机变量 X 的密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{其它} \end{cases}$$

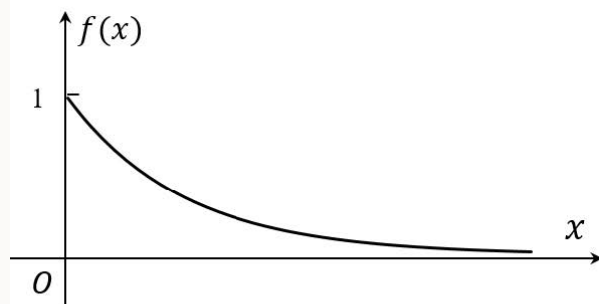
称 X 服从参数 λ 的指数分布, 记为 $X \sim e(\lambda)$.

指数分布的分布函数

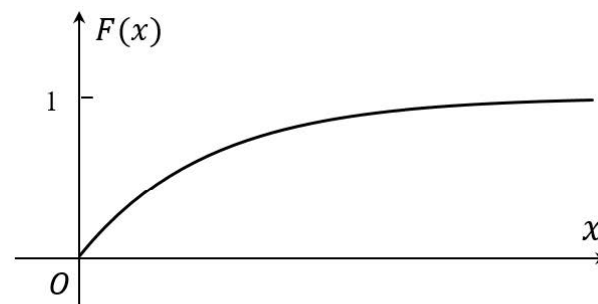
若随机变量 $X \sim e(\lambda)$ 的分布函数为

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

随机变量 $X \sim e(\lambda)$ 的密度函数和分布函数的示意图为:



密度函数



分布函数

指数分布的期望与方差

若随机变量 $X \sim e(\lambda)$, 则

$$\mathbb{E}(X) = \frac{1}{\lambda} \quad \text{VAR}(X) = \frac{1}{\lambda^2}$$

指数分布的无记忆性

定理 0.12 若随机变量 $X \sim e(\lambda)$, 则对任意 $s > 0, t > 0$, 有

$$P(X > s + t \mid X > t) = P(X > s)$$

Remarks:

- 指数分布是唯一具有无记忆性的连续型随机变量.
- 联想到：离散型随机变量 – 几何分布也是无记忆性的

$$P(X > m + n \mid X > m) = P(X > n)$$

指数分布：例 0.70

例 0.70 证明：若随机变量 X_1, \dots, X_n 相互独立的、且分别服从参数为 $\lambda_1, \dots, \lambda_n$ 的指数分布, 则有

$$X = \min\{X_1, \dots, X_n\} \sim e(\lambda_1 + \lambda_2 + \dots + \lambda_n)$$

解答：例 0.70

题目：证明：若随机变量 X_1, \dots, X_n 相互独立的、且分别服从参数为 $\lambda_1, \dots, \lambda_n$ 的指数分布, 则有

$$X = \min\{X_1, \dots, X_n\} \sim e(\lambda_1 + \lambda_2 + \dots + \lambda_n)$$

解答：

- 随机变量 X_1, \dots, X_n 相互独立的可以理解为随机变量取不同值的随机事件相互独立.
- 随机变量 X 的分布函数为

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(\min(X_1, \dots, X_n) \leq x) = 1 - P(\min(X_1, \dots, X_n) > x) \\ &= 1 - \prod_{i=1}^n P(X_i > x) = 1 - \prod_{i=1}^n \exp(-\lambda_i x) = 1 - \exp\left(-x \sum_{i=1}^n \lambda_i\right) \end{aligned}$$

由此证明完毕.

常用连续型随机变量 – 正态分布

正态分布是概率统计中最重要的一种分布, 最早由法国数学家棣莫弗在 1730s 提出, 用于近似抛硬币试验中随机事件的概率.

高斯在 1800s 首次将正态分布应用于预测天文学中星体的位置, 由此才展示出正态分布的应用价值, 正态分布因此被称为高斯分布.

正态分布的重要性主要体现在以下三个方面:

- 现实生活中很多随机现象可用正态分布进行描述, 如人的身高等;
- 很多分布可以通过正态分布来进行近似计算;
- 数理统计中常用的统计分布都是由正态分布导出的, 如 χ^2 分布、 t -分布和 F -分布.

正态分布的定义

定义 0.33 给定 $u \in (-\infty, +\infty)$ 和 $\sigma > 0$, 若连续随机变量 X 的密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x \in (-\infty, +\infty)$$

称 X 服从参数 μ, σ^2 的**正态分布**, 记为 $X \sim \mathcal{N}(\mu, \sigma^2)$.

特别地, 若 $\mu = 0$ 和 $\sigma = 1$, 称 $\mathcal{N}(0, 1)$ 为标准正态分布, 密度函数

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad x \in (-\infty, +\infty)$$

正态分布的期望和方差

若 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则

$$\mathbb{E}(X) = \mu \quad \mathbb{V}\text{AR}(X) = \sigma^2$$

若 $X \sim \mathcal{N}(0, 1)$, 则

$$\mathbb{E}(X) = 0 \quad \mathbb{V}\text{AR}(X) = 1$$

标准正态分布与一般分布的相互转换:

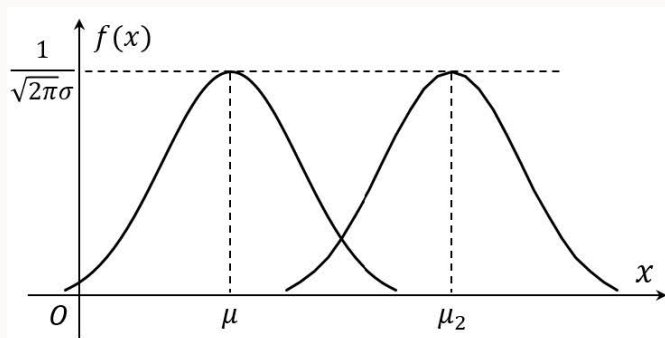
若 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则

$$Y = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

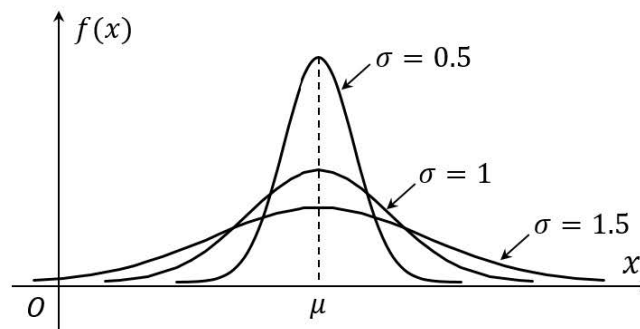
若 $X \sim \mathcal{N}(0, 1)$, 则

$$Y = \sigma X + \mu \sim N(\mu, \sigma^2)$$

正态分布的性质



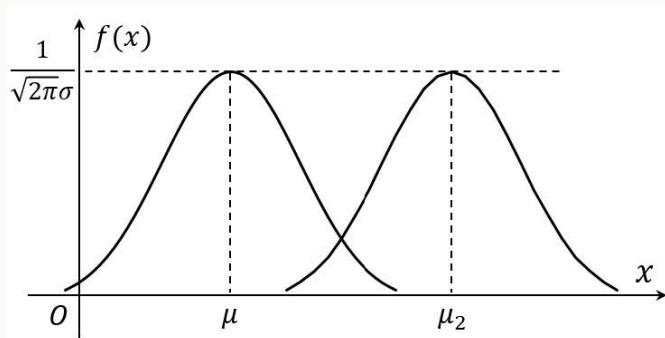
(a)



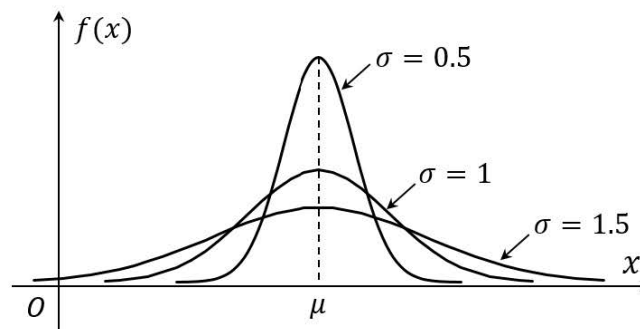
(b)

- 如图 (a) 所示, 曲线 $f(x)$ 关于直线 $x = \mu$ 对称, 即 $f(\mu - x) = f(\mu + x)$; 当 $x = \mu$ 时函数 $f(x)$ 取最大值 $f(\mu) = 1/(\sqrt{2\pi}\sigma)$;
- 如图 (a) 所示, 根据 $\lim_{x \rightarrow +\infty} f(x) = 0$ 可得曲线 $f(x)$ 的渐近线为 $y = 0$; 根据正态分布密度函数的二阶导数 $f''(x) = 0$, 可得其拐点为 $x = \mu \pm \sigma$;

正态分布的性质



(a)



(b)

- 如图 (a) 所示, 固定标准差 σ 而改变期望 μ 的值, 曲线 $f(x)$ 形状不变, 仅沿 x 轴左右平行移动;
- 如图 (b) 所示, 当 μ 固定改变 σ 的值, 根据 $f(x)$ 最大值 $f(\mu) = 1/\sqrt{2\pi}\sigma$ 可知: 当 σ 越小, 图形越陡, X 落入 μ 附近概率越大; 反之 σ 越大, 图形越平坦, X 落入 μ 附近的概率越小.

正态分布的估计：思考题

定理 0.13 若 $X \sim \mathcal{N}(0, 1)$, 对任意 $\epsilon > 0$ 有

$$P(X \geq \epsilon) \leq \frac{1}{2}e^{-\epsilon^2/2}$$

$$P(|X| \geq \epsilon) \leq \min \left(1, \sqrt{\frac{2}{\pi}} \frac{1}{\epsilon} e^{-\frac{\epsilon^2}{2}} \right)$$

在上面的定理中, 第一个不等式具有广泛的应用, 在 $\epsilon \in (0, 1)$ 时对真实的概率有更好的估计; 第二个不等式被称为 **Mill** 不等式, 在 $\epsilon \in (1, +\infty)$ 时对真实的概率有更好的估计.

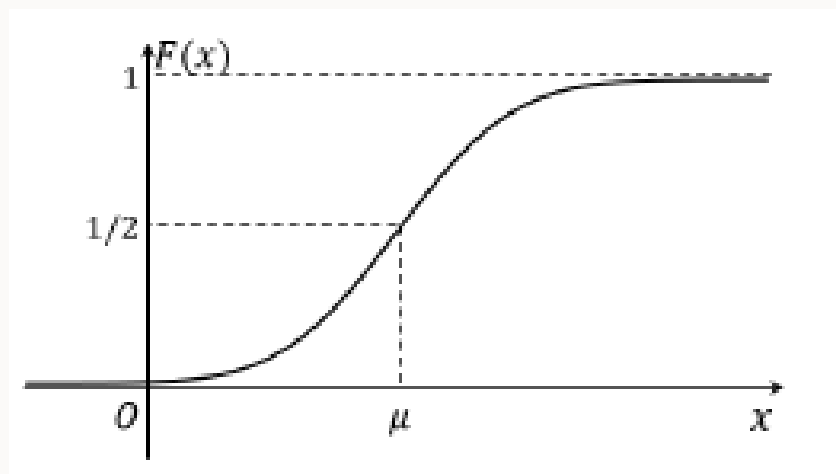
这两个不等式都可以通过定义的放缩求得.

正态分布的分布函数

若随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的分布函数为

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

该分布函数没有显示的表达式, 只能求数值解.



正态分布的分布函数

若随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的分布函数为

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

为了便于研究正态分布函数, 可将一般正态分布转化为标准正态分布 $\mathcal{N}(0, 1)$,

- 设 $X \sim \mathcal{N}(0, 1)$ 且其分布函数为:

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

- 若 $X \sim \mathcal{N}(0, 1)$, 则 $Y = \sigma X + \mu \sim N(\mu, \sigma^2)$ 满足一般正态分布

下表给出标准正态分布 $\Phi(x)$ 的函数表, 在计算具体概率时可供查询.

附：标准正态分布表 P91

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

标准正态分布函数的性质

- 根据对称性, 有 $\Phi(x) + \Phi(-x) = 1$;
- 若随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则对任意实数 $a < b$ 有

$$P(X < a) = P\left(\frac{X - \mu}{\sigma} \leq \frac{a - \mu}{\sigma}\right) = \Phi\left(\frac{a - \mu}{\sigma}\right)$$

$$P(X > b) = 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = 1 - \Phi\left(\frac{b - \mu}{\sigma}\right)$$

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right)$$

- 形成由 μ 和 $k\sigma$ 划分的概率空间

标准正态分布函数的性质

- 若随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$, 则对任意实数 $k > 0$ 有

$$P(|x - \mu| < k\sigma) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1,$$

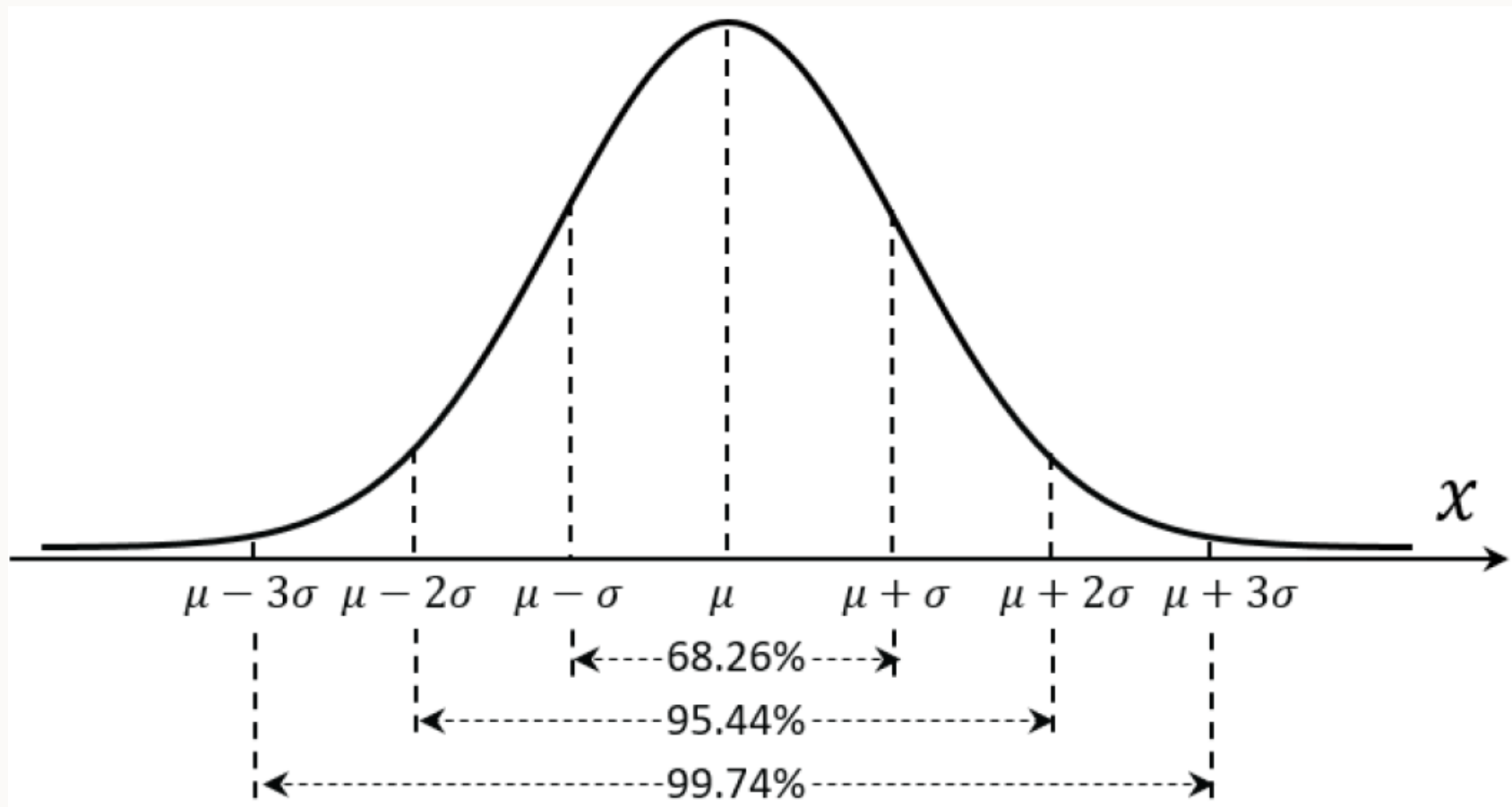
特别的, 当 $k = 1, 2, 3$ 时, 通过查标准正态分布的函数表可知:

$$P(|x - \mu| < \sigma) = 0.6826$$

$$P(|x - \mu| < 2\sigma) = 0.9544$$

$$P(|x - \mu| < 3\sigma) = 0.9974$$

随机变量 $X \sim N(\mu, \sigma^2)$ 的取值落在 $[\mu - 3\sigma, \mu + 3\sigma]$ 之外的概率不超过千分之三, 即 X 的取值几乎总在 $[\mu - 3\sigma, \mu + 3\sigma]$ 之内, 这就是人们所说的 3σ 原则, 在实际的统计推断, 特别是产品质量检测中具有重要的应用.



标准正态分布函数的性质

- 若随机变量 $X \sim \mathcal{N}(\mu, \sigma^2)$, 且已知 $P(X < c) = p$, 则有

$$p = P(X < c) = P\left(\frac{X - \mu}{\sigma} < \frac{c - \mu}{\sigma}\right) = \Phi\left(\frac{c - \mu}{\sigma}\right)$$

由此可反解出 $c = \mu + \sigma\Phi^{-1}(p)$.

- 这里 $\Phi^{-1}(x)$ 表示标准正态分布函数 $\Phi(x)$ 的反函数
- 可根据表由里向外查得, 例如 $\Phi^{-1}(0.5871) = 0.22$.

正态分布：例 0.71

例 0.71 已知某公司员工每个月的工资服从正态分布 $\mathcal{N}(6000, \sigma^2)$, 试求:

- i) 若已知标准差 $\sigma = 500$, 求工资在 5000 与 7000 之间的员工在公司中占比多少?
- ii) 当标准差 σ 为何值时, 工资在 5000 与 7000 之间的员工在公司中占比为 0.803?

解答：例 0.71

题目：已知某公司员工每个月的工资服从正态分布 $\mathcal{N}(6000, \sigma^2)$ ，试求：

- i) 若已知标准差 $\sigma = 500$ ，求工资在 5000 与 7000 之间的员工在公司中占比多少？
- ii) 当标准差 σ 为何值时，工资在 5000 与 7000 之间的员工在公司中占比为 0.803？

解答：

- 将一般正态分布转换成标准正态分布后查表求解。
- 设随机变量 X 表示公司员工每个月的工资，则 $X \sim \mathcal{N}(6000, \sigma^2)$ 。针对问题 i)，当 $\sigma = 500$ 时通过查询标准正态分布函数表有

$$P(5000 \leq X \leq 7000) = P\left(-2 \leq \frac{X - 6000}{500} \leq 2\right) = \Phi(2) - \Phi(-2) = 2\Phi(2) - 1 = 0.9544.$$

针对问题 ii)，有

$$P(5000 \leq X \leq 7000) = P\left(-\frac{1000}{\sigma} \leq \frac{X - 6000}{500} \leq \frac{1000}{\sigma}\right) = 2\Phi\left(\frac{1000}{\sigma}\right) - 1 = 0.803.$$

于是得到 $\Phi(1000/\sigma) = 0.9015$ ，通过查表可知 $\sigma \approx 775.2$ 。