术语

- outcome
- event
- probability：在一个事件中得到XX结果
    - $$P(E) = \frac{m}{N}.$$
    - ▶ **Basic:**

      (1) $P(E) \geq 0$

      (2) $P(\Omega) = P(\omega_1 \cup \omega_2 \cup \cdots \cup \omega_n) = 1$

      (3) $P(E \cup F) = P(E) + P(F)$ if E and F are ***mutually exclusive***

      ▶ **Useful:**

      Inclusion-Exclusion: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- rv 随机变量
    - discrete
        - Discrete Probability Distribution（见下文）
    - continuous

互斥&独立

If two events A and B are statistically **independent**,

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

$$P(A \cap B) = P(B) \times P(A)$$

Independence is different from ***mutually exclusive* (disjoint)** events where $P(A \cap B) = 0$

条件概率

基本

- The probability of an event (A) occurring conditional (or given) that the event B occurs is:

$$P(A|B) = P(A \cap B) / P(B)$$

where $P(B) \neq 0$

- Note that $$Pr(B) = Pr(B|A)Pr(A) + Pr(B|\bar{A})Pr(\bar{A})$$

贝叶斯

1.5 Bayes Rule

- Useful for computing conditional probability based on P(B|A) or P(A|B)

$$P(B|A) = \frac{P(A|B) \times P(B)}{P(A)}$$

- If P(A) is unknown

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)}$$

where B^c denotes "the complement of B" or "not B"

Discrete Probability Distribution

- 分布列

mean& sd

- mean-μ
- sd-σ

2.2 Mean 期望 and Standard Deviation 方差 of a Discrete RV

- We are often interested in the value we expect to arise from a random variable.
  - We call this **the expected value** = a weighted average of the possible outcomes.

$$\mu = E(X) = \sum x \cdot P(X = x)$$

- We are often interested in the variability in the values of a random variable.
  - Described using **Variance** and **Standard deviation**

$$\sigma^2 = Var(X) = E\left[\left(x - E(X)\right)^2\right] = \sum_x \left(x - E(X)\right)^2 P(X = x)$$

$$\sigma = SD(X) = \sqrt{\mathrm{Var}(X)}$$

pmf& cdf

- cdf=概率分布函数=累积分布函数
- pmf是点 cdf是面积

2.3 Cumulative Distribution Function of a Discrete RV

对 r.v. X，称 x 的函数

$$F(x) = \mathbb{P}(X \le x), \quad x \in R,$$

为 X 的**概率分布函数**或**累积分布函数** (cumulative distribution function, CDF)，简称为**分布函数**.

If X is a **discrete r.v.** with probability mass function (PMF)

$$p_k = \mathbb{P}(X = x_k), \quad k = 1, 2, \ldots,$$

or

X	x_1	x_2	x_3	$\cdots$
$\mathbb{P}$	p_1	p_2	p_3	$\cdots$

Then the cumulative distribution function (CDF) of X is

$$F(x) = \mathbb{P}(X \le x) = \mathbb{P}\Big(\bigcup_{j: x_j \le x} \{X = x_j\}\Big) = \sum_{j: x_j \le x} p_j.$$

Binomial Distribution& Bernoulli trial

- Bernoulli trial: 是在同样的条件下重复地、相互独立地进行的一种随机试验， 其特点是该随机试验只有两种可能结果：发生或者不发生。eg抛了很多次硬币
- 伯努利试验的结果加和是二项分布

- $$X \sim \text{Binom}(n, p)$$
  $$P(X = k|n, p) = f(k|n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

mean& sd

$$E(X) = E\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} E(Y_i) = \sum_{i=1}^{n} p = np$$

$$Var(X) = Var\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} Var(Y_i) = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

Multinomial Distribution 多项分布

- 该随机试验有多种可能结果eg抛骰子
- (假设六种分类 $\sum_{i=1}^{6} p_i = 1$)：
  $$P(x_1, x2, \cdots, x_6) = \frac{n!}{x_1!x_2!\cdots x_6!} p_1^{x_1} p_2^{x_2} \cdots p_6^{x_6}$$

Poisson Distribution

- eg顾客来店

  2.6 Common Discrete R.V. with Poisson Distribution 26

  在一个特定时间内，某件事情会在任意时刻随即发生，且每一次发生都是独立的

  当我们把时间段分称非常非常小的时间片构成时，在每个**时间片**内，该事件可能发生，可能不发生

  特定时间段被分称 n 个**时间片**，当时间片很小时，时间段内发生事件的概率 p 成比例减小

  但该事件在指定时间段内发生的频度相同， $n * p = \mu$ 为常数。

  营业时间 T 内有 k 个顾客到达超市的概率为： $P = \lim_{n\to\infty} C_n^k\, p^k (1-p)^{n-k}$ 且 $p = \frac{\mu}{n}$

  $$P = \lim_{n\to\infty} C_n^k \left(\frac{\mu}{n}\right)^k \left(1 - \frac{\mu}{n}\right)^{n-k} = \frac{\mu^k}{k!} e^{-\mu} = P(X = k)$$

  通常在泊松分布里， μ 被写成 λ, 某特定时间内发生的频数。

- 清华大学统计学研究中心
- 基本假设与二项分布相似

- 当没有分母（？？？无限时间）时

mean& sd

For the Poisson distribution: mean = variance = λ

eg

Examples:

王老师下课以后通常情况下有4个同学问问题， $\lambda = 4$

某一周，课程特别难，下课以后，有7个同学问问题。 $X = 7$

那么请问，7个同学问问题的可能性有多大？

$$P(X=7) = \frac{4^7}{7!}e^{-4} \approx 0.06$$

Continuous Random Variable

pdf& cdf

- pmf无了因为单点=0

3.2 Probability density function

- Every continuous random variable X has a probability density function (pdf), denoted by $f_X(x)$.
- PDF is a function such that

a $f_X(x) \geq 0$ for any $x \in \mathbb{R}$

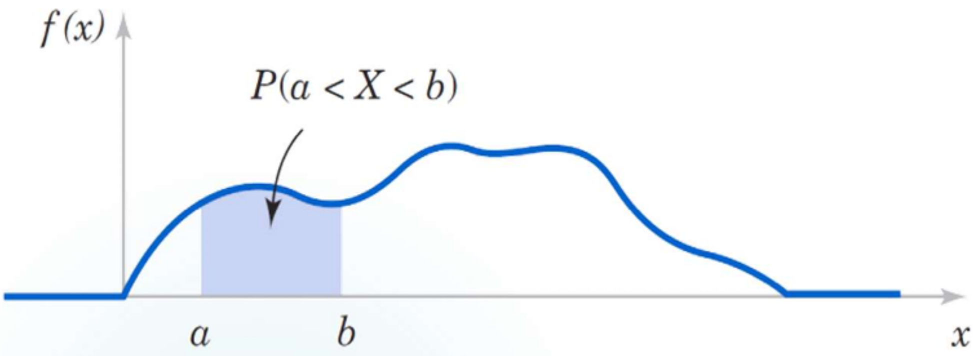
b $\int_{-\infty}^{\infty} f_X(x)dx = 1$

c $P(a \leq X \leq b) = \int_a^b f_X(x)dx$, which represents the area under $f_X(x)$ from a to b for any $b > a$.

d If x_0 is a specific value, then $P(X = x_0) = 0$. We assign 0 to area under a point.

-

3.3 Cumulative Distribution Function



Let X_0 be a specific value of interest, the **cumulative distribution function (CDF)** is defined via

$$F_X(x_0) = P(X \leq x_0) = \int_{-\infty}^{x_0} f_X(x)dx.$$

-
- $P(a < X \leq b) = \int_a^b f(t)\,dt = F(b) - F(a)$
-

mean& sd

- The mean (expectation) and variance can also be defined for a continuous random variable. Integration replaces summation in the discrete definitions.

- For a continuous random variable X. The mean of X is defined as

$$\mathrm{E}(X) = \mu_X = \int_{-\infty}^{\infty} x f_X(x) dx$$

- Let X be a continuous random variable with pdf $f_X(x)$. Suppose that g is a real-valued function. Then, g(X) is a random variable and

$$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

3.6 Variance of a Continuous R.V.

- The variance of X, denoted as Var(X) or σ^2 is

$$\sigma^2 = \mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\, dx.$$

- Standard deviation

$$\sigma = \sqrt{\sigma^2}$$

- The computational formula for variance is the same as the discrete case

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2.$$

Normal Distribution

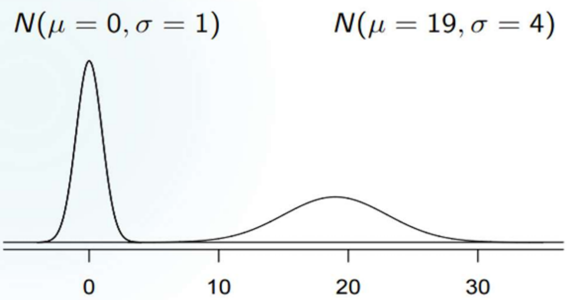
- Mean = Median = Mode
- Area under curve = 1
- Denoted as $N(\mu, \sigma^2)$ - Normal distributed with mean μ and variance σ^2

3.9.1 Normal Distribution: Formula

- The normal probability distribution is:

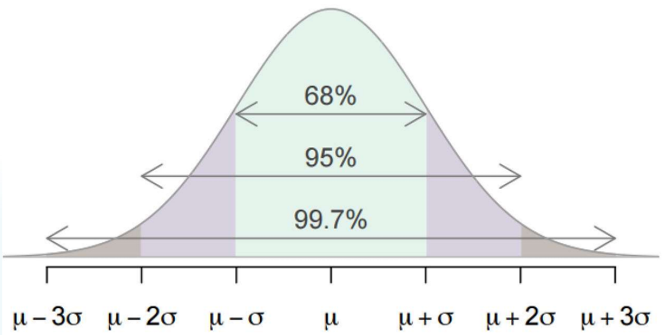
$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < +\infty$$

- $\pi \approx 3.14$ and $e \approx 2.72$
- μ, σ are mean and standard deviation parameter of the distribution



3.9.1 68-95-99.7 Rule

46



For nearly normally distributed data,
- about 68% falls within 1 SD of the mean,
- about 95% falls within 2 SD of the mean,
- about 99.7% falls within 3 SD of the mean.

Standardizing

- 为了比较

Standard Normal标准正态分布

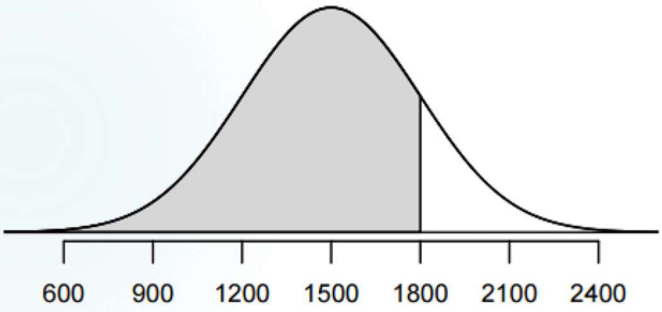
- 标准化得到的

- $Z \sim N(0,1)$
- If $X \sim N(\mu, \sigma)$, then $\frac{X-\mu}{\sigma} \sim N(0,1)$

Percentiles =Probability

左侧面积

- Percentiles is the percentage of observations that fall below a given data point.
- Graphically, percentiles is the area below the probability distribution curve to the left of that observation.



cdf

3.9.1 CDF for Standard Normal Distribution

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

Since $\varphi(t)$ is symmetric, $\Phi(x) + \Phi(-x) = 1$, or $\Phi(-x) = 1 - \Phi(x), x \in R.$

If X ~ N(μ, σ^2),

$$\begin{aligned}\mathbb{P}(X \leq a) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(a-\mu)/\sigma} \mathrm{e}^{-y^2/2}\,dy \\ &= \Phi\left(\frac{a-\mu}{\sigma}\right).\end{aligned}$$

Chi-square distribution

- 加和

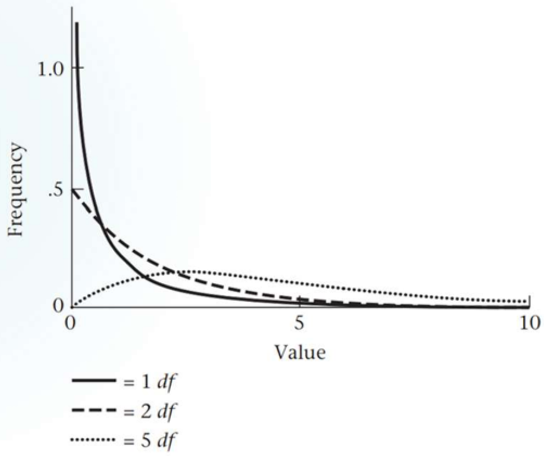
- 自由度df=n

If $G = \sum_{i=1}^{n} X_i^2$

where $X_1, \ldots, X_n \sim N(0,1)$

and the $X_i{}'s$ are independent, then G is said to follow a **chi-square distribution with n degrees of freedom (df)**. The distribution is often denoted by χ_n^2.

General shape of various χ^2 distributions with *d df*



清华大学统计学研究中心

summary

Comparison： Continuous and Discrete R.V

Continuous Random Variable

X can take on all possible values in an interval of real numbers. e.g. $X \in [0,1]$

Probability density function, $f(x)$

Cumulative distribution function, $F(x) = P(X \le x) = \int_{\infty}^{x} f(u)du$

$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$

$$\sigma^2 = V(X) = E(X-\mu)^2$$
$$= \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$
$$= E(X^2) - [E(X)]^2$$
$$= \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$$

Discrete Random Variable

X can take on only distinct 'discrete' values in a set. e.g. $X \in \{0,1,2,3,\ldots,\infty\}$

Probability mass function, $f(x)$

Cumulative distribution function, $F(x) = P(X \le x) = \sum_{x_i \le x} f(x_i)$

$\mu = E(X) = \sum_x x f(x)$

$$\sigma^2 = V(X) = E(X-\mu)^2$$
$$= \sum_x (x-\mu)^2 f(x)$$
$$= E(X^2) - [E(X)]^2$$
$$= \sum_x x^2 f(x) - \mu^2$$

清华大学统计学研究中心

3.8 Properties of Expected Value – for both discrete and continuous RV

Constant $E(c) = c$ if c is constant

Constant Multiplication $E(cX) = cE(x)$

Constant Addition $E(X + c) = E(X) + c$

Addition $E(X + Y) = E(X) + E(Y)$

Subtraction $E(X - Y) = E(X) - E(Y)$

Multiplication $E(XY) = E(X)E(Y)$ if X and Y are independent.

3.8 Properties of Variance – for both discrete anc continuous RV

Constant $Var(c) = 0$ if c is constant

Constant Multiplication $Var(cX) = c^2\ Var(x)$

Constant Addition $Var(X + c) = Var(X)$

Addition $Var(X + Y) = Var(X) + Var(Y)$ if X and Y are independent.

Subtraction $Var(X - Y) = Var(X) + Var(Y)$ if X and Y are independent.

方差可以相加