Half-time Review of STA2001

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

Outline

1. Basic Concepts of Probability Theory
2. Univariate Random Variable
3. Typical Univariate Random Distributions

Outline

1. Basic Concepts of Probability Theory
2. Univariate Random Variable
3. Typical Univariate Random Distributions

What is probability theory?

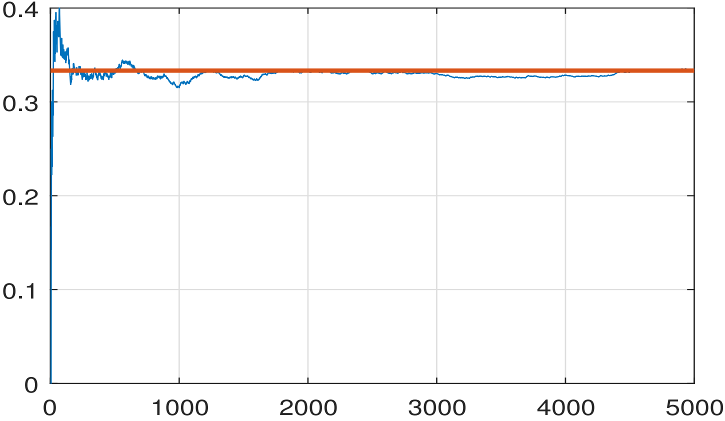
Probability theory is the branch of mathematics concerned with probability, the analysis of random phenomena - wikipedia

1. Random experiment
2. Sample space
3. Event
4. **Event A has happened**

A first definition of probability

The limit of relative frequency:

$$P(A) = \lim_{n \to \infty} \frac{\mathcal{N}(A)}{n}$$



A more formal definition of probability

Probability function is a function that assigns $P(A)$ to an event A, $A \subseteq S$ such that the following three conditions are satisfied

1. $P(A) \geq 0$
2. $P(S) = 1$
3. $A_1, A_2, \cdots$ mutually exclusive and exhaustive

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$$

Properties for probability function

- $P(A) = 1 - P(A')$
- $P(\emptyset) = 0$
- $P(A) \leq 1$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

How to calculate probability of an event

For random experiments that satisfy

Assumption 1: S contains m possible outcomes

$$e_k, \quad k = 1, 2, \cdots, m, \quad i.e., \quad S = \{e_1, e_2, \cdots, e_m\}.$$

Assumption 2: The m outcomes are “equally likely”

$$P(\{e_k\}) = \frac{1}{m}, \quad k = 1, \cdots, m.$$

$$P(A) = \frac{\mathcal{N}(A)}{\mathcal{N}(S)},$$

where $\mathcal{N}(X)$ is the number of outcomes in $X \subseteq S$.

Counting techniques — tools to calculate probability

Multiplication principle:

$$\rightarrow \boxed{\text{Experiment } E_1} \rightarrow n_1 \text{ outcomes}$$

$$\rightarrow \boxed{\text{Experiment } E_2} \rightarrow n_2 \text{ outcomes}$$

$$\rightarrow \underbrace{\boxed{\text{Experiment } E_1} \rightarrow \boxed{\text{Experiment } E_2}}_{\text{sequential implementation}} \rightarrow n_1 n_2 \text{ possible outcomes}$$

- permutation
- combination
- distinguishable permutation

Conditional Probability

Conditional probability of an event A, given that event B has occurred, is defined by

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

provided that $P(B) > 0$.

- The sample space shrinks to B.
- Conditional probability is a probability function.
- $P(A \cap B) = P(A)P(B|A)$.

Independent Events

Events A and B are independent if

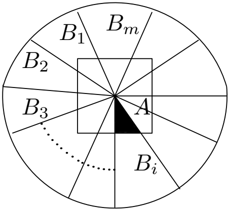
$$P(A \cap B) = P(A)P(B).$$

The occurrence of one of them does NOT change the probability of the occurrence of the other.

- A and B are independent, if and only if any pair of the following events are independent
  (a) A and B'
  (b) A' and B
  (c) A' and B'

- A, B, C are independent if
  1. pairwise independent
  2. $P(A \cap B \cap C) = P(A)P(B)P(C)$

Bayes' Theorem



Assume

1. $S = B_1 \cup B_2 \cup \cdots B_m, \quad B_i \cap B_j = \Phi$
2. $P(B_i) > 0$

Then

$$P(A) = \sum_{k=1}^{m} P(A \cap B_i) = \sum_{k=1}^{m} P(B_i) P(A|B_i)$$

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)}, \text{ provided } P(A) > 0$$

Outline

Random variable

Definition[Random Variable]

Given a random experiment with sample space S, a function $X : S \to \overline{S} \subseteq R$ that assign one real number $X(s) = x$ to each $s \in S$ is called Random Variable (RV).

Note: repeat a random experiment $\Leftrightarrow$ generate a number from $\overline{S}$

Depending on the property of $\overline{S}$

1. Discrete RV, if $\overline{S}$ is finite or countably infinite.
2. Continuous RV, if $\overline{S}$ is union of intervals.

Discrete RV Vs. Continuous RV

RV X is a function $X : S \to \overline{S} \subseteq R$

Discrete RV:

pmf $f(x) : \overline{S} \to (0,\ 1]$

1. $f(x) > 0$
2. $\sum_{x \in \overline{S}} f(x) = 1$
3. $P(X \in A) = \sum_{x \in A} f(x)$

Continuous RV:

pdf $f(x) : \overline{S} \to (0, \infty)$

1. $f(x) > 0$
2. $\int_{\overline{S}} f(x)dx = 1$
3. $P(X \in A) = \int_A f(x)dx$

Very often, we extend the definition domain of $f(x)$ from $\overline{S}$ to R by letting $f(x) = 0$ for $x \notin \overline{S}$.

Discrete RV Vs. Continuous RV

RV X is a function $X : S \to \overline{S} \subseteq R$

Discrete RV:

pmf $f(x) : R \to [0,\ 1]$

1. $f(x) \geq 0$
2. $\sum_{x \in \overline{S}} f(x) = 1$
3. $P(X \in A) = \sum_{x \in A} f(x)$

Continuous RV:

pdf $f(x) : R \to [0, \infty)$

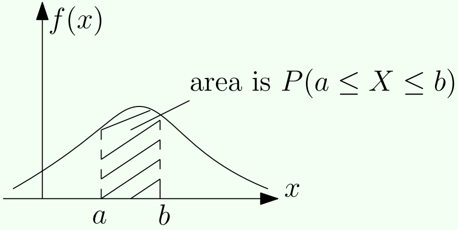
1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(X \in A) = \int_A f(x)dx$

with $f(x) = 0$ for $x \notin \overline{S}$; $\overline{S}$ is called the support set of $f(x)$.

Interpretation of pdf

Interpretation

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$



Cdf and its properties

Definition

cdf $F(x) : R \to [0, 1]$

$$F(x) = P(X \leq x), \quad \text{nondecreasing function!}$$

1. relation between the probability function and the cdf

$$P(a < X \leq b) = F(b) - F(a)$$

2. for continuous RV,

$$f(x) = F'(x)$$

for those values of x at which $F(x)$ is differentiable.

Mathematical Expectation

Definition[Mathematical Expectation]

Let X be a RV with range $\overline{S}$ and $f(x)$ be its pmf or pdf. The mathematical expectation of $g(X)$, if exists, is denoted by

$$E[g(X)] = \begin{cases} \sum_{x \in \overline{S}} g(x)f(x), & \text{discrete RV} \\ \int_{x \in \overline{S}} g(x)f(x)dx, & \text{continuous RV} \end{cases}$$

Property[Mathematical Expectation]

Mathematical expectation is a linear operator, i.e.,

$$E[c_1 g_1(X) + c_2 g_2(X)] = c_1 E[g_1(X)] + c_2 E[g_2(X)]$$

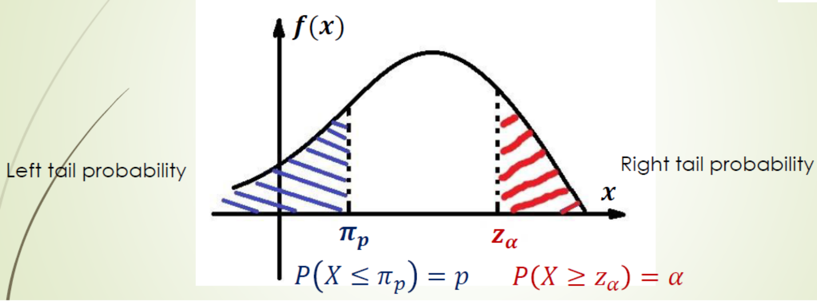
Special mathematical expectation

Definition[Special mathematical expectation]

$$E[g(X)] = \begin{cases} \sum_{x \in \overline{S}} g(x)f(x), & \text{discrete RV} \\ \int_{x \in \overline{S}} g(x)f(x)dx, & \text{continuous RV} \end{cases}$$

$$g(X) = \begin{cases} X \rightarrow \text{ Mean} \\ (X - E[X])^2 \rightarrow \text{ Variance, } Var(X) = E[X^2] - (E[X])^2 \\ X^r \rightarrow \text{ Moment} \\ e^{tX}, \text{ for } |t| < h, \rightarrow \text{ Mgf:} M^{(r)}(0) = E[X^r] \end{cases}$$

(100p)th percentile, the upper 100α percent point



Definition [(100p)th percentile]

The number π_p such that $P(X \leq \pi_p) = p$. $\pi_{0.5}, \pi_{025}$ and $\pi_{0.75}$ are called the median (the second quantile), the first and third quantiles, respectively.

Definition [the upper (100α) percent point]

The number z_α such that $P(X \geq z_\alpha) = \alpha$.

Outline

1. Basic Concepts of Probability Theory
2. Univariate Random Variable
3. Typical Univariate Random Distributions

Table of distributions

	pmf/pdf	mgf	mean	variance
Binomial ($n = 1$, Bernoulli)	$\binom{n}{x} p^x (1-p)^{n-x}$ $x = 0,1,2,\dots,n$	$[(1-p)+pe^t]^n$ $-\infty < t < \infty$	np	$np(1-p)$
Negative Binomial	$\binom{x-1}{r-1} p^r (1-p)^{x-r}$ $x = r, r+1, r+2, \dots$	$\frac{(pe^t)^r}{[1-(1-p)e^t]^r}$ $(1-p)e^t < 1$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Poisson	$\frac{\lambda^x e^{-\lambda}}{x!}, x = 0,1,2,\dots$	$e^{\lambda(e^t-1)}$	λ	λ
Gamma ($\alpha = 1$, Expnential) ($\theta = 2, \alpha = r/2, \chi^2$)	$\frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-\frac{x}{\theta}}$ $0 < x < \infty$	$\frac{1}{(1-\theta t)^\alpha}$ $t < \frac{1}{\theta}$	$\alpha\theta$	$\alpha\theta^2$
Normal ($\mu = 0, \sigma^2 = 1$, standard normal)	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2)$ $-\infty < x < \infty$	$\exp(\mu t + \frac{1}{2}\sigma^2 t^2)$ $-\infty < t < \infty$	μ	σ^2

Descriptions of distributions

	Random Phenomena
Binomial ($n = 1$, Bernoulli)	The total number of successes in n Bernoulli trials (the order does not matter)
Negative Binomial	For a given natural number r, the number of Bernoulli trials on which the rth success is observed
Poisson	The number of occurrences that a particular event happens in a given time interval or for a given physical object that can be described by APP
Gamma ($\alpha = 1$, Expnential) ($\theta = 2, \alpha = r/2, \chi^2$)	The waiting time until the αth occurrences of a particular event for an APP
Normal ($\mu = 0, \sigma^2 = 1$, standard normal)	When a large number of outcomes are observed, the outcomes have a "bell-shaped" relative frequency distribution.

Univariate normal distribution

Definition

A continuous RV X is said to be normal or Gaussian if it has a pdf of the form.

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

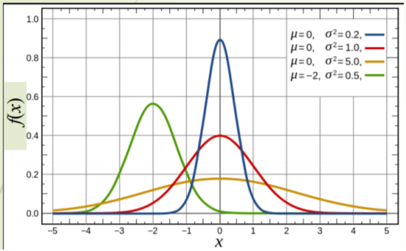
where μ and σ^2 are two parameters characterizing the normal distribution. Briefly, $X \sim N(\mu, \sigma^2)$

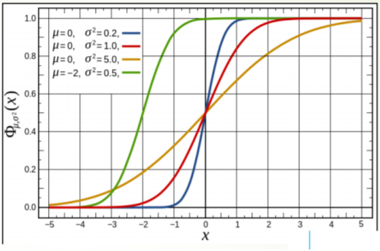
$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) dx = 1$$

Univariate normal distribution

1. Mgf:
$$M(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right), \quad t \in R$$
2. Y is said to be a standard normal distribution if
$$Y \sim N(0,1) \Leftrightarrow f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$
3. for $Y \sim N(0,1)$, $P(a \le Y \le b) = \Phi(b) - \Phi(a)$
4. if $X \sim N(\mu, \sigma^2)$, $(X-\mu)/\sigma \sim N(0,1)$ and
$$\begin{aligned} P(a \le X \le b) &= P\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$
5. if $Y \sim N(0,1)$, $Y^2 \sim \chi^2(1)$

Univariate normal distribution





$$P(Z \le z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\, dw$$

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

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026

The end

- Arrnagement of the midterm exam
- **Any questions?**

Motivations for the double integral

To prove

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(s-\mu)^2}{\sigma^2}\right) ds = 1$$

we need to handle double integrals!

Proof of $\int_{-\infty}^{\infty} f(s)ds = 1$

Let

$$I = \int_{-\infty}^{\infty} f(s)ds = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(s-\mu)^2}{\sigma^2}\right) ds \stackrel{?}{=} 1$$

Take a coordinate change $x = \frac{s-\mu}{\sigma}$

$$I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \stackrel{?}{=} 1.$$

Since $I > 0$, then if $I^2 = 1$, then $I = 1$.

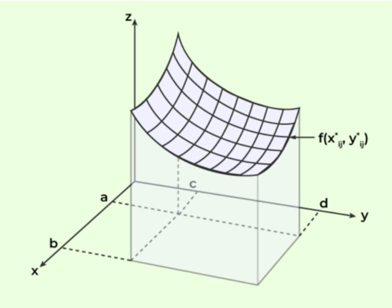
$$I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dxdy$$

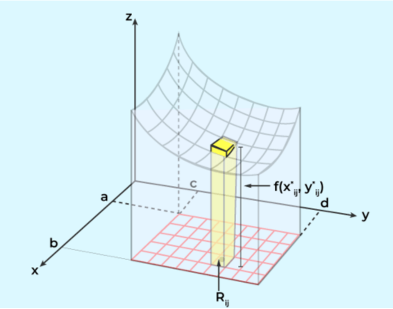
Geometric interpretation of double integral

For $f(x, y) \geq 0$

$$\text{Volume} = \iint_A f(x, y)dxdy$$

The double integral calculates the volume under the surface $f(x, y)$ over the definition domain of $f(x, y)$ or any region of interest A.





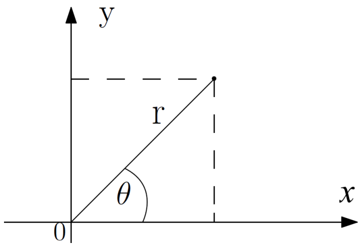
Geometric interpretation of double integral in polar coordinate system

Take a coordinate change

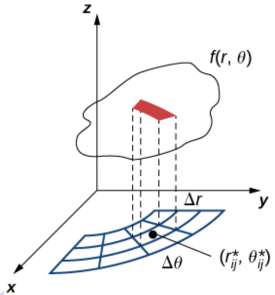
$$x = r\cos\theta$$

$$y = r\sin\theta$$

(polar coordinate)



The double integral calculates the volume under the surface $f(x, y)$ over the definition domain of $f(x, y)$ or any region of interest.



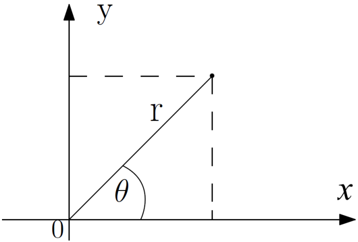
Geometric interpretation of double integral in polar coordinate system

Take a coordinate change

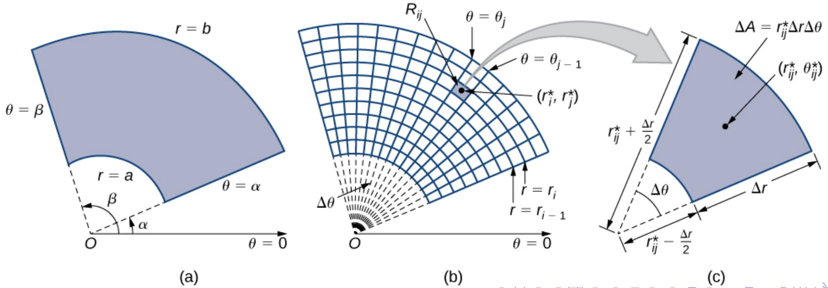
$$x = r\cos\theta$$

$$y = r\sin\theta$$

(polar coordinate)



The problem lies in to how to calculate the area of small trapezoid.



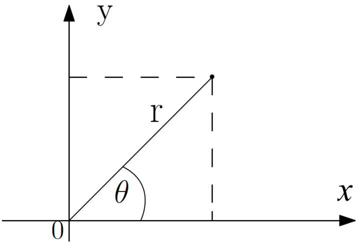
Proof of $\int_{-\infty}^{\infty} f(s)ds = 1$

Take a coordinate change

$$x = r\cos\theta$$

$$y = r\sin\theta$$

(polar coordinate)



$$I^2 = \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{-\frac{r^2}{2}} r dr d\theta = \frac{1}{2\pi}\int_0^{2\pi} d\theta \int_0^{\infty} e^{-\frac{r^2}{2}} d\frac{r^2}{2}$$
$$= \frac{1}{2\pi}\cdot 2\pi \cdot (-1)\cdot e^{-\frac{r^2}{2}}\Big|_0^{\infty} = 1$$