

Prerequisite knowledge for data science

Kunhui Zhang

March 22, 2017

1 Calculus

Logarithm function $\log(x)$, exponential function $\exp(x)$, and polynomial function $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x^1 + a_0 x^0$.

1.1 Property

Exponential function

- $e^{a+b} = e^a e^b$
- $e^{a-b} = e^a / e^b$
- $\ln_e e = 1$
- $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

Logarithm function

- $\ln(ab) = \ln(a) + \ln(b)$
- $\ln(a/b) = \ln(a) - \ln(b)$
- $\ln(1) = 0$

Polynomial function

Solve equations.

1.2 Limit Review

A limit is what happens to a function when the input approaches, but does not necessarily reach, a certain value. The general notation for a limit is:

$$\lim_{x \rightarrow c} f(x) = L$$

Rules

We set an example to review.

$$f(x) = xe^{-x} + \frac{3x}{x^2 + 1} + 2 \quad (1)$$

From above, let's find the limit for this function as x goes to ∞ .

Here we introduce two ways to think about this question. First, xe^{-x} can be rewrite as x/e^x . Note that e^x grows faster than x , and $x^2 + 1$ also increases faster than $3x$. So the answer should be 2. Another way is just use the L'Hospital's Rule, we differentiate the numerator and differentiate the denominator and then take the limit. After that, we can still get the same answer. (Write on board about the procedure)

And if we divide the function into two part, we have the nice rules below: (Adib, 2017)

Assume $\lim_{x \rightarrow c} f(x) = L$ and $\lim_{x \rightarrow c} g(x) = M$.

1. Addition:

$$\lim_{x \rightarrow c} [f(x) + g(x)] = \lim_{x \rightarrow c} f(x) + \lim_{x \rightarrow c} g(x) = L + M$$

2. Subtraction:

$$\lim_{x \rightarrow c} [f(x) - g(x)] = \lim_{x \rightarrow c} f(x) - \lim_{x \rightarrow c} g(x) = L - M$$

3. Multiplication:

$$\lim_{x \rightarrow c} [f(x) * g(x)] = \lim_{x \rightarrow c} f(x) * \lim_{x \rightarrow c} g(x) = L * M$$

4. Division:

$$\lim_{x \rightarrow c} [f(x)/g(x)] = \lim_{x \rightarrow c} f(x) / \lim_{x \rightarrow c} g(x) = L/M \quad M \neq 0$$

5. Power

$$\lim_{x \rightarrow c} f(x)^n = \left(\lim_{x \rightarrow c} f(x) \right)^n$$

6. Constant

$$\lim_{x \rightarrow c} k = k$$

7. Identity

$$\lim_{x \rightarrow c} x = c$$

1.3 Derivation

Still same example, let's take the derivation.

$$\begin{aligned} f'(x) &= x e^{-x} + \frac{3x}{x^2 + 1} + 2 \\ &= -x e^{-x} + e^{-x} + \frac{3(x^2 + 1) - 6x^2}{(x^2 + 1)^2} \\ &= (1 - x) e^{-x} + \frac{3(1 - x^2)}{(x^2 + 1)^2} \end{aligned}$$

1.4 Integral

Integral of common function

Take an example for a function $\int_2^5 (x^2 + 8x - 5) dx$

$$\begin{aligned} &\int_2^5 (x^2 + 8x - 5) dx \\ &= \left. \frac{1}{3} x^3 + 4x^2 - 5x \right|_2^5 \\ &= 108 \end{aligned}$$

Integral by part

Formula: $\int u dv = uv - \int v du$

Example: $\int 3xe^{-x}dx$

$$\begin{aligned} & \int 3xe^{-x}dx \\ &= \int 3xd - e^{-x} \\ &= -3xe^{-x} - \int -e^{-x}d(3x) \\ &= -3xe^{-x} - \int -3e^{-x}dx \\ &= -3xe^{-x} - 3e^{-x} + C \end{aligned}$$

1.5 Optimal

If $f(x)$ is differentiable and convex/concave function, then the solution to $f'(x) = 0$ is the min/max point to $f(x)$.

Eg: $x^2(e^x + 1) + 2$ First, find $f'(x)$

$$\begin{aligned} f'(x) &= (x^2(e^x + 1) + 2)' \\ &= x^2e^x + 2x(e^x + 1) \end{aligned}$$

Set $f'(x) = 0$. Note that the only possible solution for $x^2e^x + 2x(e^x + 1) = 0$ is $x = 0$. Thus, $x = 0$ is the minimum point for $f(x)$ with minimum value 2.

2 Probability

2.1 Conditional Probability

For event A and B, if $P(B) \neq 0$ the conditional probability of any event A given B is

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

In other words, $P(A|B)$ is the probability of event A after observing the occurrence of event B.

Independent

If event A and B are independent, then $P(A|B) = P(A)$ or $P(A \cap B) = P(A)P(B)$, which means B does not affect the probability of A.

Bayes Rule

Formula:

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

Eg((Zheng, 2009)): You might be interested in finding out a patient's probability of having liver disease if they are an alcoholic. "Being an alcoholic" is the test (kind of like a litmus test) for liver disease.

A could mean the event "Patient has liver disease." Past data tells you that 10% of patients entering your clinic have liver disease. $P(A) = 0.10$.

B could mean the litmus test that "Patient is an alcoholic." Five percent of the clinic's patients are alcoholics. $P(B) = 0.05$.

You might also know that among those patients diagnosed with liver disease, 7% are alcoholics. This is your $B|A$: the probability that a patient is alcoholic, given that they have liver disease, is 7%. Bayes' theorem tells you: $P(A|B) = (0.07 * 0.1)/0.05 = 0.14$

2.2 One random variable

(Maleki & Do, 2017)

Discrete random variables

- $X \sim \text{Bernoli}(p)$ (where $0 \leq p \leq 1$): one if a coin with heads probability p comes up heads, zero otherwise.

$$p(x) = \begin{cases} p & \text{if } p = 1 \\ 1 - p & \text{if } p = 0 \end{cases}$$

- $X \sim \text{Binomial}(n, p)$ (where $0 \leq p \leq 1$): the number of heads in n independent flips of a coin with heads probability p .

$$p(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

- $X \sim \text{Geometric}(p)$ (where $p > 0$): the number of flips of a coin with heads probability p until the first heads.:

$$p(x) = p(1 - p)^{x-1}$$

- $X \sim \text{Poisson}(\lambda)$ (where $\lambda > 0$): a probability distribution over the nonnegative integers used for modeling the frequency of rare events. Eg((Khare, 2015)): The number of calls coming per minute into a hotels reservation center is Poisson random variable with mean 3, find the probability for next day 5 calls. $\lambda = 3$, average number of call a day; $x = 5$, the number of calls next day.

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

Continuous random variables

- $X \sim \text{Unif}(a, b)$ (where $a < b$): equal probability density to every value between a and b on the real line. A volcano erupts once every 10 days. What is the probability you will see it erupt within 5 days? ($\int_0^5 1/10 dx$)

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{if otherwise} \end{cases}$$

- $X \sim \text{Exp}(\lambda)$ (where $\lambda > 0$): Generally the exponential distribution describes waiting time between Poisson occurrences. Eg: If jobs arrive every 15 seconds on average, $\lambda = 4$ per minute, what is the probability of waiting less than or equal to 30 seconds, i.e 0.5 min? ($\int_0^{0.5} 4e^{-4t} dt$)

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

- $X \sim \text{Norm}(\mu, \sigma^2)$ Eg: the height of 95% of students at UW are between 1.0m and 1.6m.

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

Expectation and Variance

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

Eg(Continuous case): Calculate the mean and the variance of the uniform random variable X with PDF $f_X(x) = 1, \exists x \in [0, 1], 0$ otherwise.

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x dx = \frac{1}{2}$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 dx = \frac{1}{3}$$

$$Var(X) = E[X^2] - E[X]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

Eg(Discrete case): Cat raise. 5 students raise no cat, 2 raise 1, and 1 raises 2 cats.

$$E[X] = 5/8 * 0 + 1/4 * 1 + 1/8 * 2 = 0.5$$

$$E[X^2] = 5/8 * 0 + 1/4 * 1^2 + 1/8 * 2^2 = 1.25$$

$$Var(X) = E[X^2] - E[X]^2 = 1.25 - 0.25 = 1$$

2.3 Two random variable

Joint and Marginal Distribution

The formula for joint cumulative distribution function of X and Y:

$$F_{XY}(x, y) = P(X \leq x, Y \leq y)$$

And the marginal distribution is:

$$F_X(x) = \lim_{y \rightarrow \infty} F_{XY}(x, y)$$

$$F_Y(y) = \lim_{x \rightarrow \infty} F_{XY}(x, y)$$

Joint and marginal probability mass functions

The formula for joint probability mass function of X and Y:

$$p_{XY}(x, y) = P(X = x, Y = y)$$

And the marginal probability mass functions is:

$$p_X(x) = \sum_y p_{XY}(x, y)$$

$$p_Y(y) = \sum_x p_{XY}(x, y)$$

Joint and marginal probability density functions

The formula for joint probability mass function of X and Y:

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}$$

And the marginal probability mass functions is:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx$$

Conditional distribution

Conditional distributions describes the probability distribution over Y, when we know that X must take on a certain value x. In the discrete case, when $p_X(x) \neq 0$, the conditional probability mass function of X given Y is

$$p_{Y|X}(y|x) = \frac{p_{XY}(x, y)}{p_X(x)}$$

In the continuous case, when $f_X(x) \neq 0$, analogy to discrete case, the conditional probability density function of X given Y is

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

Independence

Two random variables X and Y are independent if $F_{XY}(x, y) = F_X(x)F_Y(y)$ for all values of x and y. In discrete case, $p_{XY}(x, y) = p_X(x)p_Y(y)$; in continuous case, $f_{XY}(x, y) = f_X(x)f_Y(y)$.

Eg((He, 2011)): Given

$$f_{XY}(x, y) = \begin{cases} xy^2e^{-y} & \text{if } 0 < y < \infty, 0 < x < 1 \\ 0 & \text{if otherwise} \end{cases}$$

Determine whether X and Y are independent.

We first need to find the marginal distribution $f_X(x)$ and $f_Y(y)$.

$$\begin{aligned} f_X(x) &= \int_0^{\infty} f_{XY}(x, y) dy \\ &= x \int_0^{\infty} y^2 e^{-y} dy \\ &= x \int_0^{\infty} -y^2 de^{-y} \\ &= x(-y^2|_0^{\infty} + 2 \int_0^{\infty} ye^{-y} dy) \\ &= 2x, \quad 0 < x < 1 \end{aligned}$$

Similarly, we can find that

$$f_Y(y) = \int_0^1 f_{XY}(x, y) dx = \frac{y^2}{2} e^{-y}, \quad 0 < y < \infty$$

So, we can get:

$$f_{XY}(x, y) = f_X(x) * f_Y(y)$$

, which means X and Y are independent.

Expectation and covariance

Suppose that we have two random variables X , Y and $g: R^2 \rightarrow R$ is a function of these two random variables. Then the expected value of g is defined in the following way,

In discrete case:

$$E[g(X, Y)] = \sum_x \sum_y g(x, y) p_{XY}(x, y)$$

In continuous case:

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y) dx dy$$

And covariance is

$$Cov(X, Y) = E[(X - EX)(Y - EY)] = E[XY] - E[X]E[Y]$$

When $Cov[X, Y] = 0$, then X and Y are uncorrelated.

References

- Adib, D. (2017). Math calculus review chsn review project. <https://www.ocf.berkeley.edu/~daradib/chsntech/review/math/review-calc.pdf>.
- He, Y. (2011). Chap3. <http://www.ee.ryerson.ca/~courses/ee8103/chap3.pdf>.
- Khare, K. (2015). Lecture 15. <http://www.stat.ufl.edu/~abhisheksaha/sta4321/lect15.pdf>.
- Maleki, A. & Do, T. (2017). Review of probability theory. cs229.stanford.edu/section/cs229-prob.pdf.
- Zheng, T. (2009). Bayes' formula. <http://www.math.cornell.edu/~mec/2008-2009/TianyiZheng/Bayes.html>.