MM 225: AI and Data Science

Continuous conditional probability, Expected value, and, Variance

M P Gururajan and Hina A Gokhale

August 31, 2023

Lecture 8: Continuous conditional probability, Expected value, and, Variance

Outline

Continuous conditional probability

Discrete to continuous

- All our discrete conditional probability results: can be carried over to continuous with appropriate modifications
- Probability mass function $\rightarrow$ Probability density function
- If X is a continuous random variable with density function $f(x)$, and if E is an event with positive probability, conditional density $f(x|E)$ is defined as follows:

$$f(x|E) = \begin{cases} \frac{f(x)}{P(E)} & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases}$$

- For any event F, $P(F|E) = \int_F f(x|E)dx$ is the conditional probability of F given E

Independent events

- E and F: two events with positive probability in a continuous sample space
- E and F are independent if $P(E|F) = P(E)$; $P(F|E) = P(F)$; $P(E \cap F) = P(E)P(F)$

Definition

Definition

$\mathbf{X} = (X_1, X_2, \cdots, X_n)$ where X_i are continuous random variables associated with an experiment. The joint cumulative distribution function of $\mathbf{X}$ is defined by
$F(x_1, x_2, \cdots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n)$

Remark

$f(x_1, x_2, \cdots, x_n) = \frac{\partial^n F(x_1, x_2, \cdots, x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}$

Or,

Remark

$F(x_1, x_2, \cdots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(t_1, t_2, \cdots, t_n) dt_n dt_{n-1} \cdots dt_1$

Definition

Definition

Let $X_1, X_2 \cdots X_n$ be continuous random variables with cumulative distribution functions $F_1(x), F_2(x), \cdots F_n(x)$. Then, if for any choice of $x_1, x_2, \cdots x_n$, $F(x_1, x_2, \cdots x_n) = F(x_1)F(x_2)\cdots F(x_n)$, then these random variables are mutually independent.

Theorems

Theorem

Let $X_1, X_2 \cdots X_n$ be continuous random variables with density functions $f_1(x), f_2(x), \cdots f_n(x)$. Then, for any choice of $x_1, x_2, \cdots x_n$, $f(x_1, x_2, \cdots x_n) = f(x_1)f(x_2)\cdots f(x_n)$ is the necessary and sufficient condition for these random variables to be mutually independent.

Theorem

Let $X_1, X_2 \cdots X_n$ be mutually independent continuous random variables and let $\phi_1(x)$, $\phi_2(x)$, $\cdots$, $\phi_n(x)$ be continuous functions. Then, $\phi_1(X_1)$, $\phi_2(X_2)$, $\cdots$, $\phi_n(X_n)$ are mutually independent.

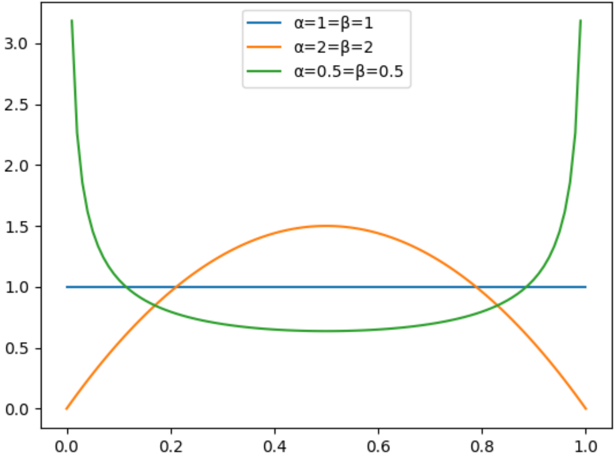
A sequence $X_1, X_2, \cdots X_n$ of random variables X_i that are mutually independent and have the same density is called an independent trials process.
Recall *iid*: independent, identical distributions!

Examples

Problem of drug effectiveness

- Assumption: a drug is effective with probability x each time it is used
- Assumption: various trials are independent
- x: a continuous random variable between zero and unity
- What density function describes the distribution for x?
- Assume beta density for x before experiment
- Beta density: sample space a combination of discrete and continuous coordinates
- $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$ where α, β are integers
- $$B(\alpha, \beta, x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Beta: density function



Experiment

- x before experiment; described by beta density
- Give drugs to n subjects and record the number of successes; say i
- i: discrete random variable
- Set of possible outcomes of experiment: ordered pair (x, i)
- $m(i|x) = \binom{n}{i} x^i (1-x)^{n-i}$
- $f(x, i) = m(i|x)B(\alpha, \beta, x)$
- If i successes are observed, the probability density that the drug is effective is $B(\alpha + i, \beta + n - i)$; if the drug is now tested on the next subject, the probability that the drug is effective is $\frac{\alpha+i}{\alpha+\beta+n}$
- For large n, probability of success after experiment $\approx$ the proportion of successes in the experiment
- Suppose there are two drugs: how to choose between them?

Two-armed bandit problem

Suppose in a Casino, there are two slot machines. Each machine either gives 1 dollar or nothing. The probabilities for the first and second machines to give 1 dollar are x and y, respectively. Suppose you play the game ten times, each time choosing one or the other machine. Suppose you want to maximize the number of times you win, what should be the strategy of play?

Strategies

- **Play the best machine**
  At every stage, calculate the probability of pay off for each machine and choose the machine with the highest probability.
- Weakness: chance might make you continue playing with the machine with lower return
- **Play the winner**
  Play the same machine when you win but switch machine when you lose
- **Homework:** Simulate the strategies
- **Importance:** Two armed bandit: important in clinical trials!

Expected value

A couple of problems

- Suppose a die is rolled. If odd number turns up, G wins an amount equal to the number. If even number turns up, G loses an amount equal to the number. Is this a reasonable game for G to play?
- Let us generate a number from the interval [0,1]. If you generate a large number of such random numbers, what is the average value of the generated numbers?

Definitions

Definition

Let X be a discrete random variable with sample space Ω and distribution function $m(x)$. The expected value $E(X) = \sum_{x \in \Omega} xm(x)$ provided this sum converges absolutely. We typically refer to $E(X)$ as the mean and denote it by μ. X does not have an expected value if there is no convergence.

Definition

Let X be a real valued random variable with density function $f(x)$. The expected value $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$ provided the integral $\int_{-\infty}^{\infty} |x|f(x)dx$ is finite.

Cauchy distribution

- Consider the random variable X with density $f_X(x) = \frac{a}{\pi} \frac{1}{a^2+x^2}$
- No expectation: integral diverges
- No variance either!

Properties

Theorem

Let $X_1, X_2, \cdots X_n$ be random variables with finite expected values (discrete or continuous). Let $c_1, c_2, \cdots c_n$ are constants. Then,
$E(c_1X_1 + c_2X_2 + \cdots c_nX_n) = c_1E(X_1) + c_2E(X_2) + \cdots c_nE(X_n)$.

Remark

First fundamental mystery of probability: Expectations add irrespective of whether the summands are mutually independent or not.

Theorems

Theorem

If X is a discrete random variable with sample space Ω and distribution function $m(x)$, and if $\phi : \Omega \to R$ is a function, then, $E(\phi(X)) = \sum_{x \in \Omega} \phi(x) m(x)$ provided the series converges absolutely.

Theorem

If X is a real-valued random variable and if $\phi : \Omega \to R$ is a continuous real-valued function with domain [a,b], then, $E(\phi(X)) = \int_{-\infty}^{\infty} \phi(x) f_X(x) dx$ provided the integral exists.

Independence

Theorem

If $X_1, X_2, \cdots X_n$ are n mutually independent random variables (discrete or continuous) with expectation values $E(X_i)$, then, $E(X_1X_2\cdots X_n) = E(X_1)E(X_2)\cdots E(X_n)$

A problem: Let $Z = (X, Y)$ be a point chosen at random in the unit square. Let $A = X^2$ and $B = Y^2$. What is $E(AB)$?

- A and B are independent. Why? If random variables are independent, their functions are also independent.
- So, the answer is easy using the above theorem! $(\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9})$.
- Calculation from definition: Complicated. $f(AB) = -\frac{\log(t)}{4\sqrt{t}}$

Conditional expectation

Definition

If F is any event and X is a random variable with sample space $\Omega = \{x_1, x_2, \cdots\}$, then the conditional expectation given F is $E(X|F) = \sum_j x_j P(X = x_j|F)$.

Theorem

Let X be a random variable with sample space Ω. If $F_1, F_2, \cdots F_r$ are events such that $F_i \cap F_j = \emptyset$ if $i \neq j$ and $\Omega = \cup_j F_j$, then, $E(X) = \sum_j E(X|Fj)P(F_j)$.

Martingales

Coin toss problem

- **Recall the problem:**
  Suppose we toss a coin. Assume the coin is fair. If head (H) comes up, G(uru) gets Re. 1 and H(ina) loses Re. 1; if tail (T) comes up, H gets Re. 1 and G loses Re. 1. Suppose the coin is tossed 40 times.
- Let $S_1, S_2, \cdots S_n$ be G's accumulated fortune in playing this game.
- $E(S_n | S_{n-1} = a, \cdots, S_1 = r) = \frac{1}{2}(a+1) + \frac{1}{2}(a-1) = a$
- G's expected fortune in the next play is equal to his present fortune: game is *fair*
- Martingale: a fair game
- If the coin is not fair, $E(S_n | S_{n-1} = a, \cdots, S_1 = r) = p(a+1) + q(a-1) = a + p - q$
- If $p < q$ $(p > q)$, game is unfavourable (favourable)
- Can we make a fair game, a favourable game? No
- Martingale doubling system : tutorial applying this to stock market
- Efforts to make unfavourable game, favourable!

Variance

Definitions

Definition

Let X be a discrete random variable with expected value $E(X) = \mu$. Then, the variance of X is $V(X) = E((X-\mu)^2)$. Further, $V(x) = \sum_x (x-\mu)^2 m(x) = E(X^2) - \mu^2$ where m is the distribution function of X.

Definition

Let X be a real valued random variable with density function $f(x)$. The variance $\sigma^2 = V(X) = E((X-\mu)^2)$. Further, $V(x) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = E(X^2) - \mu^2$ where $\mu = E(X)$.

Remark

Standard deviation: $\sigma = \sqrt{V(X)}$

Properties

- If X is any random variable, and c is any constant, $V(cX) = c^2 V(X)$ and $V(X + c) = V(X)$.
- If X and Y are two independent random variables, $V(X + Y) = V(X) + V(Y)$

Standardisation

Theorem

Let $X_1, X_2, \cdots X_n$ be an independent trials process with $E(X) = \mu$ and $V(x) = \sigma^2$. Let $S_n = X_1 + X_2 + \cdots + X_n$ be the sum and $A_n = \frac{S_n}{n}$ be the average. Then, $E(S_n) = n\mu$, $V(S_n) = n\sigma^2$, $E(A_n) = \mu$ and $V(A_n) = \frac{\sigma^2}{n}$.

Remark

Standard deviation of the average goes to 0 as $n \to \infty$.

Remark

If we set $S_n^\star = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$, then, $E(S_n^\star) = 0$ and $V(S_n^\star) = 1$.

THANK YOU!!!