

# 基于连续型随机变量性质的积分运算公式

许 岷

北京工业大学数学统计学与力学学院, 北京

收稿日期: 2024年7月15日; 录用日期: 2024年8月9日; 发布日期: 2024年8月16日

## 摘 要

面向工科类和经济类学生的概率论与数理统计教学实践中, 积分运算是教学的难点。本文利用连续型随机变量的性质, 推导了基于正态分布、伽马分布和贝塔分布的积分运算公式, 解决某些复杂函数的积分运算问题, 并将其应用于贝叶斯统计中。

## 关键词

连续型随机变量, 积分运算, 贝叶斯统计

# Integral Operation Formulas Based on the Properties of Continuous Random Variables

Min Xu

School of Mathematics, Statistics and Mechanics, Beijing University of Technology, Beijing

Received: Jul. 15<sup>th</sup>, 2024; accepted: Aug. 9<sup>th</sup>, 2024; published: Aug. 16<sup>th</sup>, 2024

## Abstract

In the teaching practice of probability theory and mathematical statistics for engineering and economics students, the integral operation is difficult for students. By using the properties of continuous random variables, this paper derives several common integral operation formulas based on normal, gamma and beta distributions, computes the integrals of complex functions, and applies them to Bayesian statistics.

## Keywords

Continuous Random Variables, Integral Operation, Bayesian Statistics



## 1. 引言

《概率论与数理统计》是本科教学的数学基础课，《高等数学》中微积分知识是本课程的基础。在面向工科类和经济类学生的教学实践中，针对复杂函数积分问题，如  $\int_{-\infty}^{\infty} e^{-x^2} dx$ 、 $\int_0^{\infty} x^{\alpha-1} e^{-\lambda x} dx$  ( $\alpha > 0, \lambda > 0$ ) 和  $\int_0^1 x^{a-1} (1-x)^{b-1} dx$  ( $a > 0, b > 0$ ) 等，由于积分的相关知识点掌握不牢，大多数学生存在畏惧心理。教学课时有限，教师往往对上述积分计算无法详细展开，导致积分运算问题成为了《概率论与数理统计》教学实践中的痛点和难点。

事实上，利用连续型随机变量的定义、概率密度函数及其数字特征等性质，能够推导出一系列积分运算公式，简化复杂函数的积分问题。例如，文献[1]和[2]总结了正态分布和几类反常积分的关系。文献[3]介绍了利用正态分布和指数分布计算积分  $\int_{-\infty}^{\infty} e^{-x^2} dx$  和  $\int_0^{\infty} (9x^2 + 6x + 1)e^{-3x} dx$  等复杂函数积分。文献[4]以积分  $\int_{-\infty}^{\infty} e^{-x^2/2} dx$  的计算为例，强调了利用概率论求解积分的方法应在教学中加以重视。利用这类积分运算公式，一方面简化了《高等数学》课程中涉及的复杂函数积分问题；另一方面有助于学生熟悉《概率论与数理统计》的基本知识。

与上述工作从具体的积分问题出发不同，本文基于常用连续型随机变量的性质，总结和归纳了若干积分运算公式，以《概率论与数理统计》中的典型题目为切入点，说明运用这类积分运算公式在解决复杂函数积分时的优势，为教学实践中微积分知识的讲授方式提供一种新角度。最后，本文强调了这类积分运算公式在贝叶斯统计中的一些应用。

## 2. 预备知识

本节介绍连续型随机变量的定义及其性质。

定义[5]：设  $X$  是一个随机变量， $F(x)$  是其分布函数，如果存在一个定义在  $(-\infty, \infty)$  上的非负实值函数  $f(x)$ ，使得

$$F(x) = \int_{-\infty}^x f(t) dt,$$

则称  $X$  为连续型随机变量， $f(x)$  为  $X$  的概率密度函数。

性质 1 [5]：设  $f(x)$  是某个连续型随机变量的概率密度函数，则

(i)  $f(x) \geq 0, -\infty < x < \infty$ ;

(ii)  $\int_{-\infty}^{\infty} f(x) dx = 1$ 。

性质 2 [5]：设连续型随机变量  $X$  的概率密度函数为  $f(x)$ ，若  $X$  的期望  $E(X)$  存在，则

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx.$$

利用性质 1 和性质 2，结合常用连续型随机变量的概率密度函数和期望，能够直接计算一些复杂函数的积分。

## 3. 常用连续型随机变量及其积分运算公式

本节介绍正态分布、伽马分布和贝塔分布等常用连续型分布的性质，并探讨对应的积分运算公式。

### 3.1. 正态分布

设连续型随机变量  $X$  服从均值为  $\mu$  和方差为  $\sigma^2$  的正态分布, 记为  $X \sim N(\mu, \sigma^2)$ , 概率密度函数为

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

由性质 1 和性质 2, 得到如下积分运算公式:

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

$$\int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu\sqrt{2\pi\sigma^2}. \quad (2)$$

例: 设随机变量  $X$  的分布函数为

$$F(x) = 0.5\Phi(x) + 0.5\Phi\left(\frac{x-4}{2}\right),$$

其中  $\Phi(x)$  为标准正态分布函数, 求期望  $E(X)$ 。

解: 先求  $X$  的概率密度函数为

$$f(x) = \frac{dF(x)}{dx} = \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + \frac{1}{4} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-4)^2}{8}}.$$

求  $X$  的期望为

$$E(X) = \frac{1}{2} \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + \frac{1}{4} \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-4)^2}{8}} dx,$$

这是一个复杂函数的积分, 一般需使用变量变换和伽马函数等知识, 有一定难度。然而, 利用公式 (2) 能够简化积分计算。对第一个积分, 可视为标准正态分布的期望, 设  $X_1 \sim N(0,1)$ , 则

$$\frac{1}{2} \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1}{2} E(X_1) = 0.$$

对第二个积分, 可视为正态分布  $N(4, 2^2)$  的期望, 设  $X_2 \sim N(4, 2^2)$ , 则

$$\frac{1}{4} \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-4)^2}{8}} dx = \frac{2}{4} \int_{-\infty}^{\infty} x \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-4)^2}{2 \times 2^2}} dx = \frac{1}{2} E(X_2) = 2.$$

故  $E(X) = 0 + 2 = 2$ , 此题目为 2017 年全国硕士研究生统一招生考试数学(一)试题。

利用积分运算公式还可以计算标准正态分布的  $k$  阶矩  $E(X^k)$ 。

例: 设随机变量  $X$  服从标准正态分布, 即  $X \sim N(0,1)$ , 求  $X$  的  $k$  阶矩  $E(X^k)$ 。

解: 根据期望的定义

$$E(X^k) = \int_{-\infty}^{\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

当  $k$  为奇数时, 被积函数为奇函数, 故  $E(X^k) = 0$ ; 当  $k$  为偶数时,

$$E(X^k) = \int_{-\infty}^{\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{2^{k/2}}{\sqrt{\pi}} \int_0^{\infty} t^{k-1} e^{-t} dt = (k-1)(k-3)\cdots 1.$$

利用标准正态随机变量的  $k$  阶矩和极坐标变换，能够推导出一类复杂三角函数的积分运算公式。

例：当  $k$  为偶数时，计算  $\int_0^{\pi/2} (\sin \theta)^k d\theta$ 。

解：当  $k$  为偶数时，令

$$I_k = \int_{-\infty}^{\infty} x^k \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = (k-1)(k-3)\cdots 1.$$

则有

$$I_k^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u^k v^k e^{-\frac{1}{2}(u^2+v^2)} dudv.$$

作极坐标变换  $u = r \cos \theta$ ， $v = r \sin \theta$ ，

$$I_k^2 = \frac{1}{2\pi} \int_0^{\infty} \int_0^{2\pi} r^{2k} \cos^k \theta \sin^k \theta e^{-r^2/2} r d\theta dr = \frac{1}{2\pi} \left[ \int_0^{\infty} r^{2k+1} e^{-r^2/2} dr \right] \left[ \int_0^{2\pi} \cos^k \theta \sin^k \theta d\theta \right].$$

其中第一个中括号的积分

$$\int_0^{\infty} r^{2k+1} e^{-r^2/2} dr = 2^k \int_0^{\infty} t^k e^{-t} dt = 2^k \Gamma(k+1),$$

第一个等号利用变量变换  $t = r^2/2$ ，第二个等号利用伽马函数。此时，能够计算复杂三角函数的积分

$$\int_0^{2\pi} \cos^k \theta \sin^k \theta d\theta = \frac{2\pi I_k^2}{2^k \Gamma(k+1)} = \frac{\pi(k-1)(k-3)\cdots 1}{2^{k-1} k(k-2)\cdots 2}, k = 2, 4, 6, \dots$$

如果对上述积分做适当的变量变换，有

$$\int_0^{2\pi} \cos^k \theta \sin^k \theta d\theta = \int_0^{2\pi} \left( \frac{1}{2} \sin 2\theta \right)^k d\theta = \frac{1}{2^{k-2}} \int_0^{\pi/2} (\sin t)^k dt,$$

第二个等号利用变量变换  $t = 2\theta$ 。由此，得到

$$\int_0^{\pi/2} (\sin t)^k dt = \frac{(k-1)(k-3)\cdots 1}{k(k-2)\cdots 2} \cdot \frac{\pi}{2}, k = 2, 4, 6, \dots$$

上述复杂三角函数的积分是《高等数学》课程“周期函数的定积分”小节的重点和难点，通常使用分部积分法和递归法则计算。正态随机变量推导出的积分运算公式，提供了求解这类复杂函数积分的一个新视角。

### 3.2. 伽马分布

设连续型随机变量  $X$  服从参数为  $\alpha > 0, \lambda > 0$  的伽马分布，记为  $X \sim \text{Gamma}(\alpha, \lambda)$ ，其概率密度函数为

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

其中  $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$  为伽马函数。由性质 1，得到如下积分计算公式

$$\int_0^{\infty} x^{\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha)}{\lambda^\alpha}. \tag{3}$$

例：设随机变量  $T$  服从自由度为  $n$  的  $t$  分布，求随机变量  $T$  的期望和方差。

解：在本科教学中，由于  $t$  分布的概率函数密度函数较为复杂，讲授其性质时一般只介绍  $t$  分布的期望和方差的结果，不给出计算过程。本例利用  $t$  分布的定义和公式(3)推导一种求解  $t$  分布数字特征的方法，该方法简化了积分求解过程。

若随机变量  $T$  服从自由度为  $n$  的  $t$  分布，则  $T$  可表示为

$$T = \frac{X}{\sqrt{Y/n}},$$

其中  $X \sim N(0,1)$  服从标准正态分布， $Y \sim \chi_n^2$  服从自由度为  $n$  的卡方分布，且  $X$  和  $Y$  相互独立。求期望

$$E(T) = E(XY^{-1/2}n^{1/2}) = \sqrt{n}E(X)E(Y^{-1/2}),$$

其中第二个等号是因为随机变量  $X$  和  $Y$  的独立性。因为  $E(X) = 0$ ，所以  $E(T) = 0$ 。求方差

$$\text{Var}(T) = E(T^2) - [E(T)]^2 = E(X^2Y^{-1}n) = nE(X^2)E(Y^{-1}),$$

其中最后一个等号是因为随机变量  $X$  和  $Y$  的独立性。计算

$$E(X^2) = \text{Var}(X) + [E(X)]^2 = 1$$

和

$$E(Y^{-1}) = \int_0^{\infty} y^{-1} \frac{1}{2^{n/2} \Gamma(n/2)} y^{n/2-1} e^{-y/2} dy = \frac{1}{2^{n/2} \Gamma(n/2)} \int_0^{\infty} y^{n/2-2} e^{-y/2} dy.$$

利用公式 (3) 进行积分运算，

$$\frac{1}{2^{n/2} \Gamma(n/2)} \int_0^{\infty} y^{n/2-2} e^{-y/2} dy = \frac{1}{2^{n/2} \Gamma(n/2)} \times \frac{\Gamma(n/2-1)}{(1/2)^{n/2-1}} = \frac{1}{n-2},$$

需满足  $n/2-1 > 0$ ，即  $n > 2$ 。由此，得  $\text{Var}(T) = \frac{n}{n-2}$ ， $n > 2$ 。

### 3.3. 贝塔分布

设连续型随机变量  $X$  服从参数为  $a > 0$ ， $b > 0$  的贝塔分布，记为  $X \sim \text{Beta}(a, b)$ ，其概率密度函数为

$$f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}, & 0 < x < 1, \\ 0, & \text{其他}, \end{cases}$$

由性质 1，得到如下积分计算公式

$$\int_0^1 x^{a-1} (1-x)^{b-1} dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \quad (4)$$

在《高等数学》中，公式(4)也被称为贝塔函数或  $B$  函数，对工科类和经济类学生，一般属于选学内容。利用贝塔分布导出的积分运算公式，能够简化积分求解过程。在《概率论与数理统计》中该积分应用广泛，例如求解最小次序统计量的数字特征。

例：设随机变量  $X_1, X_2, \dots, X_n$  相互独立，均服从  $(0, \theta)$  上的均匀分布，记  $Y = \min\{X_1, X_2, \dots, X_n\}$ ，求期望  $E(Y)$ 。

解：因为  $X_i \sim U(0, \theta)$ ，其概率密度函数和分布函数分别为

$$f(x) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta, \\ 0, & \text{其他,} \end{cases} \text{ 和 } F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{\theta}, & 0 \leq x < \theta, \\ 1, & x \geq \theta. \end{cases}$$

当  $0 < t < \theta$  时,  $Y$  的概率密度函数为

$$f_Y(t) = n[1 - F(t)]^{n-1} f(t) = n \left(1 - \frac{t}{\theta}\right)^{n-1} \frac{1}{\theta}.$$

求期望

$$E(Y) = \int_0^\theta n \left(1 - \frac{t}{\theta}\right)^{n-1} \frac{1}{\theta} dt = n \int_0^1 z(1-z)^{n-1} dz = n\theta \frac{\Gamma(2)\Gamma(n)}{\Gamma(n+2)} = \frac{\theta}{n+1}.$$

其中第二个等号利用变量变换  $z = t/\theta$ , 第三个等号利用公式(4)计算。

#### 4. 在贝叶斯统计中的应用

本文提出的积分运算公式在贝叶斯统计中能够简化某些复杂函数的积分运算。设  $X$  和  $\theta$  为连续型随机变量, 连续型场合下的贝叶斯公式为

$$f(\theta|x) = \frac{f(\theta)f(x|\theta)}{\int f(\theta)f(x|\theta)d\theta}.$$

其中  $f(\cdot|\cdot)$  表示条件概率密度。一般  $\theta$  为参数, 称  $f(\theta)$  为先验分布,  $f(\theta|x)$  为观测到  $x$  后的后验分布。简化上式, 得到贝叶斯统计框架: 参数后验  $f(\theta|x) \propto$  参数先验  $f(\theta) \times$  数据  $f(x|\theta)$ 。积分运算公式能够简化积分  $\int f(\theta)f(x|\theta)d\theta$  的计算, 且当  $\theta$  是随机向量时能够计算参数的边缘密度函数。

##### 4.1. 托马斯·贝叶斯的台球实验

贝叶斯公式的提出者托马斯·贝叶斯(Thomas Bayes)是位牧师, 有优秀的数学素养, 擅长通过数学思维解决实际问题[6]。为了探索因果关系, 贝叶斯牧师进行了如下试验: 假设有一个长方形的台球桌, 长为  $L$ 。贝叶斯牧师背对台球桌, 让助手随机地抛出一个球到桌面上, 记该球的水平位置为  $X$ , 重复抛球  $n$  次, 并汇报落在第一个位置  $X$  左侧的次数。随着试验次数的增加, 通过试验结果, 贝叶斯牧师能够猜测出该球可能落在的区域。

令  $\theta = X/L$ , 其中  $X$  未知。假设  $\theta$  的先验分布为  $(0,1)$  上的均匀分布。助手进行  $n$  次试验, 其中  $y$  次落在第一次的左边, 利用贝叶斯公式, 能够计算第一次球的水平位置落在区间  $[a,b]$  的概率:

$$P(a < X < b) = \frac{P(a < X < b, y)}{f(y)} = \frac{1}{f(y)} \int_{a/L}^{b/L} C_n^y \theta^y (1-\theta)^{n-y} d\theta,$$

其中  $C_n^y = \frac{n!}{y!(n-y)!}$  表示组合数,

$$f(y) = \int_0^1 C_n^y \theta^y (1-\theta)^{n-y} d\theta = \frac{n!}{y!(n-y)!} \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)} = \frac{1}{n+1},$$

第二个等号利用公式(4)计算。

##### 4.2. 正态分布假设下的后验分布

当参数  $\theta$  是随机向量时, 如正态分布的参数  $\theta = (\mu, \sigma^2)$ , 利用贝叶斯公式能够得到联合后验密度函数,

求解边缘后验密度函数的积分较为复杂，积分运算公式能够简化积分计算。本节以正态分布为例，介绍简化过程。

假设  $x_1, x_2, \dots, x_n$  是来自服从正态总体  $N(\mu, \sigma^2)$  的随机样本，记  $\tau = 1/\sigma^2$  表示分布的精度。在贝叶斯统计框架下，有

$$f(\mu, \tau | x_1, x_2, \dots, x_n) \propto f(\mu, \tau) \times f(x_1, x_2, \dots, x_n | \mu, \tau).$$

假设参数服从无信息先验分布  $f(\mu, \tau) \propto \tau$ ，参数的后验分布为

$$f(\mu, \tau | x_1, x_2, \dots, x_n) \propto \tau \times \tau^{n/2} \exp\left\{-\frac{\tau}{2}[(n-1)s^2 + n(\bar{x} - \mu)^2]\right\},$$

其中

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

首先，根据参数的联合后验密度函数求解  $\tau$  的边缘密度函数为

$$\begin{aligned} f(\tau | x_1, x_2, \dots, x_n) &\propto \int_{-\infty}^{\infty} \tau^{n/2+1} \exp\left\{-\frac{\tau}{2}[(n-1)s^2 + n(\bar{x} - \mu)^2]\right\} d\mu \\ &\propto \tau^{n/2+1} \exp\left\{-\frac{\tau(n-1)s^2}{2}\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{\tau n(\bar{x} - \mu)^2}{2}\right\} d\mu \\ &\propto \tau^{n/2+1} \exp\left\{-\frac{\tau(n-1)s^2}{2}\right\} \tau^{-1/2} = \tau^{\frac{n+1}{2}} \exp\left\{-\frac{\tau(n-1)s^2}{2}\right\}, \end{aligned}$$

其中最后一行的正比号利用公式(1)计算。显然，参数  $\tau$  的后验分布是伽马分布，即

$$\tau | x_1, x_2, \dots, x_n \sim \text{Gamma}\left(\frac{n+3}{2}, \frac{(n-1)s^2}{2}\right).$$

其次，求解  $\mu$  的边缘密度函数为

$$\begin{aligned} f(\mu | x_1, x_2, \dots, x_n) &\propto \int_0^{\infty} \tau^{n/2+1} \exp\left\{-\frac{\tau}{2}[(n-1)s^2 + n(\bar{x} - \mu)^2]\right\} d\tau \\ &\propto \frac{2^{\frac{n+4}{2}} \Gamma\left(\frac{n+4}{2}\right)}{[(n-1)s^2 + n(\bar{x} - \mu)^2]^{\frac{n+4}{2}}} \propto \left[1 + \frac{n(\bar{x} - \mu)^2}{(n-1)s^2}\right]^{-\frac{n+4}{2}}, \end{aligned}$$

其中第二行第一个正比号利用公式(3)计算。

## 5. 结论

本文基于常用的连续型随机变量的分布，如正态分布、伽马分布和贝塔分布等，推导了对应的积分运算公式。利用这些积分运算公式能够简化某些复杂函数的积分，为求解复杂积分问题提供新思路，值得在《概率论与数理统计》教学中加以重视和推广。本文还把这些积分运算公式应用在贝叶斯统计中，介绍了历史上托马斯·贝叶斯牧师的桌球实验和正态分布假设下参数的边缘密度函数的计算方法。

## 致谢

作者非常感谢相关文献对本文的启发以及审稿专家提出的宝贵意见。

## 参考文献

- [1] 张春春, 涂俐兰. 基于正态分布求解两类反常积分[J]. 高等数学研究, 2017, 20(1): 66-68.
- [2] 赵雯雪, 屈志扬, 侯文. 有关正态分布广义积分计算的一些注记[J]. 应用数学进展, 2023, 12(6): 2951-2957.
- [3] 贺慧敏, 李萍, 范钦伟. 巧用概率论方法求解积分问题[J]. 高等数学研究, 2024, 27(2): 93-95.
- [4] 王彭德, 戈芳. 考研数学试题中的概率积分[J]. 高等数学研究, 2020, 23(4): 79-81.
- [5] 王松桂, 张忠占, 程维虎, 高旅端. 概率论与数理统计[M]. 第4版. 北京: 科学出版社, 2023.
- [6] 陈希孺. 数理统计简史[M]. 哈尔滨: 哈尔滨工业出版社, 2021.