从概率到信息熵

陈童

July 5, 2025

Contents

统计物理是研究大量粒子的行为，微观上，这些粒子都满足哈密顿力学规律，而宏观上则会涌现出诸如温度和压强等新的物理量，统计物理的目标就是在微观和宏观之间建立桥梁。而这就需要在力学规律的基础上引入一些新的概念，最核心的就是概率和信息熵的概念，这也就是本章要讲述的内容。

1 概率论的一般公理

概率论是用来描述随机事件的，其背后有一个随机变量，记为x, 它有多种可能的取值，记这些不同可能性的集合为$\mathcal{S}$。$\mathcal{S}$可以是一个离散的集合，比如单个比特的两种可能状态$\{0,1\}$, 比如色子的六种可能状

态$\{1,2,3,4,5,6\}$。$\mathcal{S}$也可以是一个连续的集合，比如某个随机出现的粒子的位置坐标。

一个事件就是$\mathcal{S}$的某个子集$E \subseteq \mathcal{S}$, 只要E中的任何一个可能性得以实现，我们就称事件E发生了。设有两个事件A和B, 如果定义A发生或者B发生为一个新的事件，我们就记这个新的事件为$A \cup B$, 也就是A和B的并集。相反，如果定义A发生且B也发生为一个新的事件，那就记它为$A \cap B$, 也就是A, B的交集。如果$A \cap B = \emptyset$, 其中$\emptyset$表示空集，则称A, B为互不相容的事件。

概率论的核心就是给每一个事件E赋予一个发生的概率$P(E)$，它要满足如下三条公理：

(i) 非负性公理：即对于任意事件A, 有$P(A) \geq 0$。即概率值不能为负数。

(ii) 归一化公理: 即$P(\mathcal{S}) = 1$, 换言之必然发生的事件概率为1。

(iii) 加法公理：即对于任意两个不相容事件A, B，有$P(A \cup B) = P(A) + P(B)$。结合前两条公理，立即有一条明显的推论，即对于任何A, 均有$0 \leq P(A) \leq 1$。

对于$\mathcal{S}$为离散集合的情况，满足这些公理的最简单办法是，对每一种单独的可能性$i \in \mathcal{S}$都赋予一个概率$0 \leq p_i \leq 1$, 并且要求它们满足如下归一化条件

$$\sum_{i \in \mathcal{S}} p_i = 1. \tag{1}$$

对于$\mathcal{S}$为连续集合的情形，如果仅仅只有一个随机变量，那满足上述公理的常见办法是，对于每一个$x \in \mathcal{S}$引入一个所谓的概率密度$p(x) \geq 0$(有时也记作$\rho(x)$), 并取

$$P(A) = \int_A p(x)dx, \tag{2}$$

式中$\int_A$表示在集合A上积分。当然，$p(x)$要满足如下归一化条件

$$P(\mathcal{S}) = \int_{\mathcal{S}} p(x)dx = 1. \tag{3}$$

如果有多个随机变量$\mathbf{x} = (x_1, x_2, \cdots, x_N)$, 每一个都是连续取值的，依

然记相应的状态空间为$\mathcal{S}$, 则可以引入联合概率密度

$$p(\mathbf{x}) = p(x_1, x_2, \cdots, x_N). \tag{4}$$

它满足如下归一化条件

$$P(\mathcal{S}) = \int_{\mathcal{S}} d^N\mathbf{x} p(\mathbf{x}) = 1, \tag{5}$$

式中$d^N\mathbf{x} = \prod_{i=1}^N dx_i$。假设我们仅仅只关心这$N$个随机变量中的一部分，则相应的概率密度可以通过将其余变量积分掉来得到，比方说

$$p(x_1, \cdots, x_m) = \int \prod_{i=m+1}^{N} dx_i p(x_1, \cdots, x_N). \tag{6}$$

而，如果这N个随机变量相互独立(也称作统计独立)，则联合分布$p(\mathbf{x})$是每个变量单独分布的乘积，即

$$p(\mathbf{x}) = \prod_{i=1}^{N} p_i(x_i). \tag{7}$$

在已知事件B发生的前提下(当然要求$P(B) > 0$)，事件A发生的概率，称为事件A在事件B发生的条件下的条件概率，记作$P(A|B)$, 其定义是

$$P(A|B) \equiv \frac{P(A \cap B)}{P(B)}. \tag{8}$$

其含义就是事件$A \cap B$在事件B中所占的比例。根据条件概率的这个定义，很容易推导出如下**贝叶斯定理：**

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{9}$$

贝叶斯定理的常见应用如，将A解释为模型或假说，将B解释为观测数据。则$P(A)$就是观测之前对模型的先验信念，$P(B|A)$是基于模型或假说对数据可能性的推断，$P(B)$则是实验中具体数据出现的概率，最后，$P(A|B)$则是实验之后，数据的支撑对模型选择的依据。

如果事件B的发生与否对事件A发生的概率没有影响，反之亦然，即满足$P(A|B) = P(A)$或$P(B|A) = P(B)$, 则就称A和B是相互独立的，有时候也说是统计独立的。很显然，对于两个相互独立的事件，我们有

$$P(A \cap B) = P(A)P(B). \tag{10}$$

设$B_1, B_2, ..., B_n$是一个完备事件组。即满足：第一，B_i两两互不相容，$B_i \cap B_j = \emptyset (i \neq j)$。第二, $\cup_{i=1}^{n} B_i = \mathcal{S}$。则不难看出

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i). \tag{11}$$

例：新冠检测。设有一种常规新冠检测法的灵敏度和可靠度均为99%, 即患者(D)99%将检出为阳性(+), 而健康者(N)99%将检出为阴性($-$)。已知一群体的新冠患病率为0.5%. 问每个检出为阳性的个体真正患病的概率有多大?

先写下已知的所有概率

$$P(D) = 0.005, \quad P(+|D) = 0.99, \quad P(-|N) = 0.99. \tag{12}$$

于是，$P(N) = 1 - P(D) = 0.995$, $P(+|N) = 1 - P(-|N) = 0.01$。代入这些数据，不难得出总阳性率为

$$P(+) = P(+|D)P(D) + P(+|N)P(N) = 0.0149. \tag{13}$$

最后由贝叶斯定理，即有

$$P(D|+) = \frac{P(+|D)P(D)}{P(+)} = \frac{0.99 \times 0.005}{0.0149} = 0.332. \tag{14}$$

也即是说，尽管检测法可靠性不低，但是由于人群患病率不高，所以即使检测出阳性，个体真正患病的概率也不算高。

2 信息熵

本节介绍信息论的一些基本概念。正如下一章将会看到的，统计物理就是在哈密顿力学的基础上，通过引入这些新概念而建立起来的。

2.1 香农熵

假设我们收到一条消息，它是由字母a和b所构成的字符串，比如

$$aababbaaaab\cdots \tag{15}$$

假设字母a的出现概率为p, 字母b的出现概率为$1-p$。假设某个人接受到这样的一长串消息，比方说共N个字母，N很大，我们想问，他可以从这串消息中提取多少比特的信息？

上面这句话是什么意思呢？意思就是，在这个人读取消息的具体内容之前，他是无知的，或者说他对消息的内容是不确定的，所谓他能提取的信息量，意思就是，当他读取消息的具体内容之后，他被消除的无知有多少？或者说他被消除的不确定度有多少？

为了衡量具体消息消除掉的无知有多少，或者说消除掉的不确定性有多少，不妨数一下可能的消息共有多少条。根据概率的含义，如果N很大，那么这样的消息中大约会包含有pN个字母a, 包含有$(1-p)N$个字母b, 很显然，这样的消息的可能数目为

$$\Omega = \frac{N!}{(pN)!((1-p)N)!}. \tag{16}$$

根据斯特林公式，当n很大时，$n! \sim \sqrt{2\pi n}(n/e)^n$(或者等价的，$\ln(n!) \approx n\ln n - n + \frac{1}{2}\ln(2\pi n)$)。所以，当$N$足够大时，近似有

$$\Omega \sim \frac{N^N}{(pN)^{pN}((1-p)N)^{(1-p)N}} = \frac{1}{p^{pN}(1-p)^{(1-p)N}} = 2^{NS}, \tag{17}$$

式中(这里log**是以2为底的对数**)

$$S = -p\log p - (1-p)\log(1-p). \tag{18}$$

注意，单个比特仅有$0, 1$两种可能状态，因此也就是说，如果要把这些消息用比特记录下来，我们至少需要NS个比特，其中每一条具体消息对应这NS个比特组的一种可能状态。因此，我们就称，每一条具体消息能够消除的无知或者说消除的不确定度为NS比特，因为它使得这个比特组的状态确定了！换言之，每一条具体消息的信息量为NS比特，平均每个字母的信息量为S比特。

推而广之，假设这些字母是从一个更大的字母表$\{a_1, a_2, ..., a_k\}$中选取的，字母a_i的出现概率为p_i(不妨记这个概率分布为A)。那么当字符串很长时，即字符串有N个字母且N很大时，字母a_i大约会出现p_iN次。因此，总的可能消息的数目大约为

$$\frac{N!}{(p_1N)!(p_2N)!\cdots(p_kN)!} \sim \frac{N^N}{\prod_{i=1}^k (p_iN)^{p_iN}} = 2^{NS_A}, \tag{19}$$

式中S_A为

$$S_A = -\sum_{i=1}^{k} p_i \log p_i. \tag{20}$$

它就是平均每个字母所包含的信息量，称之为**香农熵**。

根据上面的讲述可以知道，对于一个概率分布为A的字母表，香农熵S_A就是平均每读取一个字母所能消除掉的无知量，或者说所能消除掉的不确定度。从另一个方面来说，S_A也就是了解具体字母之前，我们对每一个字母具体为何的无知量，或者说对每一个字母的不确定度。

很显然，可能的消息总量2^{NS_A}一定大于1，从而可知

$$S_A \geq 0. \tag{21}$$

要让$S_A = 0$, 那总共就只能有一条可能消息，也就是说，只可能出现一个字母，它出现的概率为1，其它字母出现的概率都为零。对于一个共k个字母的字母表，最大可能的香农熵在每个字母均以相等概率$1/k$出现的时候取到，这时候

$$S_A = -\sum_{i=1}^{k} (1/k) \log(1/k) = \log k. \tag{22}$$

读者可以利用拉格朗日乘子法证明这个结论，其中要最大化的泛函为$S(\{p_i\}) = -\sum_i p_i \log p_i$, 约束条件为$\sum_i p_i = 1$。

很显然，只要有一个概率分布就可以定义相应的香农熵，它就是这个概率分布中所包含的信息量，也就是我们了解随机变量具体的可能状态之前对它的无知量。假设记随机变量的状态空间为$\mathcal{S}$，概率分布为$p_i,\ i \in \mathcal{S}$(依然记为分布A)，则相应的香农熵就是

$$S_A \equiv -\sum_{i \in \mathcal{S}} p_i \log p_i. \tag{23}$$

当然，香农熵可以直接推广到连续的概率分布, 假设记概率密度为$p(x)$，则相应的香农熵就是，

$$S_A \equiv -\int_{\mathcal{S}} dx p(x) \log p(x). \tag{24}$$

香农熵的可加性：假设有两个独立的随机变量，它们的概率分布分别为$A=\{p_i\}$、$B=\{q_j\}$, 由于是独立的随机变量，所以这两个分布也是统计独立的。记这两个随机变量的联合分布为AB，很显然$AB=\{p_iq_j\}$。则由于

$$\begin{aligned}-\sum_{i,j}p_iq_j\log(p_iq_j)&=-\sum_{i,j}p_iq_j(\log p_i+\log q_j)\\&=-\sum_i p_i\log p_i-\sum_j q_j\log q_j.\end{aligned}\tag{25}$$

这就说明

$$S_{AB}=S_A+S_B.\tag{26}$$

推广到连续随机变量也是类似的。总之，独立随机变量的香农熵有可加性。

由于香农熵是具体消息所能消除的无知量，而如果把消息擦除了，那无知当然就增加了，所以，**擦除信息会导致香农熵增加！**

2.2 相对熵

假设某个随机变量X，其状态空间为$\mathcal{S}=\{i=1,2,\cdots,s\}$, 其真实的概率分布为$p_X=\{p_i\}$, 但是我们不知道，我们猜测了一个描写它的概率分布$q_X=\{q_i\}$, 那我们这个猜测中额外包含的无知是多少呢？

为了回答这个问题，假设我们重复N次观察这个随机变量的取值，N很大，则状态i出现的次数大约为p_iN次，根据上一小节的分析，可能的观测序列共有$\frac{N!}{\prod_{j=1}^s(p_jN)!}\sim 2^{-N\sum_j p_j\log p_j}$个。但是由于我们认为这个随机变量的概率分布是$q_X$, 所以我们会认为每一个观测序列出现的概率为$\prod_{i=1}^s q_i^{p_iN}$, 所以相关观测结果出现的总概率为

$$P=\prod_{i=1}^s q_i^{p_iN}\frac{N!}{\prod_{j=1}^s(p_jN)!}\sim 2^{-N\sum_i p_i(\log p_i-\log q_i)}=2^{-NS(p_X\|q_X)}.\tag{27}$$

式中

$$S(p_X\parallel q_X)\equiv\sum_i p_i(\log p_i-\log q_i)=\sum_i p_i\log\big(\frac{p_i}{q_i}\big).\tag{28}$$

$NS(p_X \parallel q_X)$就是我们的猜测中额外包含的无知，平均来说，我们的猜测相对于每一次观测的额外无知是$S(p_X \parallel q_X)$, 它就称之为概率分布q_X相对于概率分布p_X的**相对熵**。

由于概率P总小于等于1，所以很明显必然有

$$S(p_X \parallel q_X) \geq 0. \tag{29}$$

等于号当且仅当$q_X = p_X$时取到(相应于$P = 1$), 即仅当我们的猜测完全正确时，额外的无知才是零。值得注意的是，$S(p_X \parallel q_X)$关于p_X和q_X是不对称的。另外，相对熵的概念当然也可以推广到连续随机变量，比方说

$$S(p_X \parallel q_X) \equiv \int dx p(x) \log\big(\frac{p(x)}{q(x)}\big) \geq 0. \tag{30}$$

由于(29)非常重要，这里不妨指出另一个独立的证明。为此只需注意到

$$\log(x) - 1 + 1/x \geq 0, \tag{31}$$

等于号当且仅到$x = 1$时取到。因此

$$S(p_X \parallel q_X) = \sum_i p_i \log\big(\frac{p_i}{q_i}\big) \geq \sum_i p_i(1 - \frac{q_i}{p_i}) = 0. \tag{32}$$

互信息

假设有两个随机变量X和Y, X的状态空间为$\{x_1, \ldots, x_k\}$, Y的状态空间为$\{y_1, \cdots, y_r\}$, X, Y的联合概率分布为$p_{XY}(x_i, y_j)$, 相应的香农熵记作S_{XY}。进而可以得到X和Y各自单独的概率分布，为

$$p_X(x_i) = \sum_j p_{XY}(x_i, y_j), \quad p_Y(y_j) = \sum_i p_{XY}(x_i, y_j). \tag{33}$$

相应的香农熵分别记作S_X和S_Y。我们想知道，通过观测这两个随机变量中的一个，能够对另一个了解多少？也就是，两个随机变量的相互关联中所包含的信息量有多少？

为此我们定义**互信息**$I(X;Y)$, 它由如下相对熵给出

$$\begin{aligned} I(X;Y) &\equiv \sum_{i,j} p_{XY}(x_i,y_j)\log\big(\frac{p_{XY}(x_i,y_j)}{p_X(x_i)p_Y(y_j)}\big) \\ &= \sum_{i,j} p_{XY}(x_i,y_j)\big(\log p_{XY}(x_i,y_j) - \log p_X(x_i) - \log p_Y(y_j)\big) \\ &= S_X + S_Y - S_{XY}. \end{aligned} \tag{34}$$

最后一行我们利用了(33)式，以及各香农熵的定义。根据相对熵的非负性，很显然有

$$I(X;Y) = S_X + S_Y - S_{XY} \geq 0. \tag{35}$$

等于号当且仅当$p_{XY}(x_i,y_j) = p_X(x_i)p_Y(y_j)$, 也就是两个随机变量相互独立时，才成立。从上面这个定义可知，$I(X;Y)$衡量是根据对X和Y各自的单独观察，并猜测两者相互统计独立时，所包含的额外无知，换言之，也就是两个随机变量的相互关联中所包含的信息量。因此当这两个随机变量相互独立(不相关)时，$I(X;Y) = 0$。

3 期望值

考虑到在统计物理中的应用，本节主要以单个连续随机变量为例进行讨论。

随机变量x的任何函数$\mathcal{O}(x)$依然是随机变量，可以定义其统计期望值为

$$\langle \mathcal{O}(x) \rangle \equiv \int_{\mathcal{S}} dx p(x)\mathcal{O}(x). \tag{36}$$

特别的，$\mathcal{O}(x) = x^n$的统计期望值，称之为随机变量x的n阶矩，记作M_n

$$M_n = \langle x^n \rangle. \tag{37}$$

概率密度$p(x)$的傅里叶变换称之为随机变量x的特征函数，

$$\widetilde{p}(k) = \langle e^{-ikx} \rangle = \int dx p(x) e^{-ikx}. \tag{38}$$

反过来当然也有

$$p(x) = \frac{1}{2\pi}\int dk\widetilde{p}(k)e^{ikx}. \tag{39}$$

很显然，特征函数是n阶矩的生成函数，即

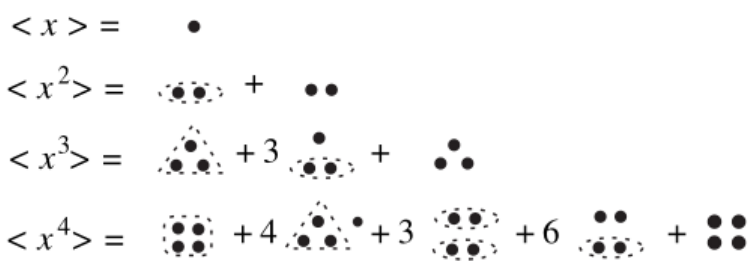
$$\widetilde{p}(k) = \langle\sum_{n=0}^{\infty}\frac{(-ik)^n}{n!}x^n\rangle = \sum_{n=0}^{\infty}\frac{(-ik)^n}{n!}\langle x^n\rangle. \tag{40}$$

也即是说，将特征函数$\widetilde{p}(k)$进行泰勒展开，展开项第n阶的系数基本上就是n阶矩。

进一步可以按照下式定义n阶累积量(Cumulants)$\langle x^n\rangle_c$

$$\widetilde{p}(k) = \exp\big(\sum_{n=0}^{\infty}\frac{(-ik)^n}{n!}\langle x^n\rangle_c\big). \tag{41}$$

一个重要的定理使得我们可以方便地根据累积量算出矩，即：首先，将n阶累积量表示成n个点的连通集团。则通过把m个点分解成更小的集团(连通或者不连通)，并对所有可能的分解求和，即可以表示m阶矩。其中，每一种可能分解的贡献由相应连通子集团所表示的累积量的乘积表示。举例如下图:



对应的代数表达式为

$$\begin{aligned}
&\langle x\rangle = \langle x\rangle_c\\
&\langle x^2\rangle = \langle x^2\rangle_c + \langle x\rangle_c^2 \Rightarrow \langle x^2\rangle_c = \langle x^2\rangle - \langle x\rangle^2\\
&\langle x^3\rangle = \langle x^3\rangle_c + 3\langle x^2\rangle_c\langle x\rangle_c + \langle x\rangle_c^3\\
&\langle x^4\rangle = \langle x^4\rangle_c + 4\langle x^3\rangle_c\langle x\rangle_c + 3\langle x^2\rangle_c^2 + 6\langle x^2\rangle_c\langle x\rangle_c^2 + \langle x\rangle_c^4.
\end{aligned}$$

这个定理的证明可以通过比较(40)式和(41)式完成

$$\sum_{n=0}^{\infty}\frac{(-ik)^n}{n!}\langle x^n\rangle=\exp\big(\sum_{n=0}^{\infty}\frac{(-ik)^n}{n!}\langle x^n\rangle_c\big)=\prod_n\sum_{l_n}\big[\frac{(-ik)^{nl_n}}{l_n!}\big(\frac{\langle x^n\rangle_c}{n!}\big)^{l_n}\big].\quad(42)$$

通过比较两边$(-ik)^m$项的系数，即可得

$$\langle x^m\rangle=\sum_{\{l_n\}}{}'\,m!\prod_n\frac{1}{l_n!(n!)^{l_n}}\langle x^n\rangle_c^{l_n},\quad(43)$$

式中的求和表示对整数m的所有满足$\sum nl_n=m$的分解求和。这个表达式中的数值因子刚好就是将m个点分解成$\{l_n\}$个n点集团的可能方式数。

4 数学附录

这个数学附录其实是为下一章作准备的，读者可以在读到下一章的相关部分再来参考。

(一)，高斯积分

高斯积分（也称为概率积分）是数学中的一个重要积分，定义为：

$$I=\int_{-\infty}^{\infty}e^{-x^2}dx$$

它的结果是$\sqrt{\pi}$，即$I=\sqrt{\pi}$。下面我将一步步推导这个结果。推导的核心技巧是利用双重积分和极坐标变换。

推导步骤：

1. 定义积分：令$I=\int_{-\infty}^{\infty}e^{-x^2}dx$。我们可以考虑它的平方来简化问题：

$$I^2=\left(\int_{-\infty}^{\infty}e^{-x^2}dx\right)\left(\int_{-\infty}^{\infty}e^{-y^2}dy\right)$$

由于两个积分独立，我们可以将它们合并为一个双重积分：

$$I^2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-(x^2+y^2)}dxdy$$

这个双重积分表示对整个xy-平面的积分。

2. 转换为极坐标：为了计算这个双重积分，我们使用极坐标变换。令：

$$x = r\cos\theta, \quad y = r\sin\theta$$

那么：

$$x^2 + y^2 = r^2, \quad \text{且} \quad dxdy = rdrd\theta$$

代入后，双重积分变为：

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-r^2} \cdot rdrd\theta$$

3. 分离变量（角度和径向部分）：由于积分可分离，我们将它拆分为两个独立积分的乘积：

$$I^2 = \left(\int_0^{2\pi} d\theta\right)\left(\int_0^{\infty} e^{-r^2} rdr\right)$$

4. 计算角度部分：角度积分很简单：

$$\int_0^{2\pi} d\theta = \theta\Big|_0^{2\pi} = 2\pi$$

5. 计算径向部分：径向积分是$\int_0^{\infty} e^{-r^2} rdr$。这是一个标准积分，我们可以通过换元法求解。令$u = r^2$，则：

$$du = 2rdr \implies rdr = \frac{du}{2}$$

积分限变化：当$r = 0$时，$u = 0$; 当$r \to \infty$ 时，$u \to \infty$。代入后：

$$\int_0^{\infty} e^{-r^2} rdr = \int_0^{\infty} e^{-u} \cdot \frac{du}{2} = \frac{1}{2}\int_0^{\infty} e^{-u} du = \frac{1}{2}.$$

6. 合并结果：代入回I^2的表达式：

$$I^2 = (2\pi) \times \left(\frac{1}{2}\right) = 2\pi \cdot \frac{1}{2} = \pi$$

7. 求解I：于是：

$$I^2 = \pi \implies I = \pm\sqrt{\pi}$$

由于被积函数e^{-x^2}总是正数，积分结果必须是正的，因此我们取正根：

$$I = \sqrt{\pi}.$$

（二），推导n维笛卡尔空间中单位球面的面积

在n 维笛卡尔空间$\mathbb{R}^n$中，单位球面定义为满足$x_1^2 + x_2^2 + \cdots + x_n^2 = 1$的所有点构成的集合。其面积（即超球面的表面积）记为Ω_n。我们将利用高斯积分来推导Ω_n的表达式。

步骤1: 回顾高斯积分

在n 维空间中，我们考虑径向对称的高斯积分：

$$I_n = \int_{\mathbb{R}^n} e^{-(x_1^2+x_2^2+\cdots+x_n^2)} dx_1 dx_2 \cdots dx_n$$

因为被积函数仅依赖于$r^2 = x_1^2 + x_2^2 + \cdots + x_n^2$，且每个变量独立，我们可以将积分分解为$n$ 个一维高斯积分的乘积：

$$I_n = \left(\int_{-\infty}^{\infty} e^{-x^2} dx \right)^n = (\sqrt{\pi})^n = \pi^{n/2}$$

所以，

$$I_n = \pi^{n/2}.$$

步骤2: 在极坐标下表达I_n

在n维空间中，使用极坐标变换更为方便。令：$r = \sqrt{x_1^2 + \cdots + x_n^2}$（径向距离）。则体积元变换为：

$$d^n\mathbf{x} = r^{n-1} dr d\Omega_n$$

其中$d\Omega_n$是单位球面上的面积元，因此单位球面的面积Ω_n定义为：

$$\Omega_n = \int d\Omega_n$$

现在，将I_n在极坐标下展开：

$$I_n = \int_{\text{所有}\mathbf{x}} e^{-r^2} d^n\mathbf{x} = \int_{\text{所有角度}} \int_0^{\infty} e^{-r^2} r^{n-1} dr d\Omega_n$$

因为被积函数独立于角度，积分可分离为：

$$I_n = \left(\int_0^{\infty} e^{-r^2} r^{n-1} dr \right) \left(\int d\Omega_n \right) = \left(\int_0^{\infty} e^{-r^2} r^{n-1} dr \right) \Omega_n.$$

令$J_n = \int_0^\infty e^{-r^2} r^{n-1} dr$，则：

$$I_n = J_n \cdot \Omega_n.$$

步骤3: 计算径向积分J_n

我们需要计算$J_n = \int_0^\infty e^{-r^2} r^{n-1} dr$。通过变量替换将其与Gamma 函数关联。令$t = r^2$，则：$r = t^{1/2}$, 且$dr = \frac{1}{2} t^{-1/2} dt$。

当$r = 0$ 时，$t = 0$；当$r \to \infty$ 时，$t \to \infty$。代入积分：

$$J_n = \int_0^\infty e^{-t} (t^{1/2})^{n-1} \cdot \frac{1}{2} t^{-1/2} dt = \int_0^\infty e^{-t} t^{(n-1)/2} \cdot \frac{1}{2} t^{-1/2} dt$$

简化指数：$t^{(n-1)/2 - 1/2} = t^{n/2-1}$，因此：

$$J_n = \frac{1}{2} \int_0^\infty e^{-t} t^{n/2-1} dt$$

Gamma 函数的定义为$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$，所以：

$$J_n = \frac{1}{2} \Gamma\left(\frac{n}{2}\right).$$

步骤4: 结合结果求解Ω_n

综合上面的结果，有：

$$I_n = \pi^{n/2} = J_n \cdot \Omega_n$$

代入J_n：

$$\pi^{n/2} = \frac{1}{2} \Gamma\left(\frac{n}{2}\right) \cdot \Omega_n$$

解得Ω_n:

$$\Omega_n = \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}.$$