1 Probability : Worked Examples

(1.1) The *information entropy* of a distribution $\{p_n\}$ is defined as $S = -\sum_n p_n \log_2 p_n$, where n ranges over all possible configurations of a given physical system and p_n is the probability of the state $|n\rangle$. If there are Ω possible states and each state is equally likely, then $S = \log_2 \Omega$, which is the usual dimensionless entropy in units of $\ln 2$.

Consider a normal deck of 52 distinct playing cards. A new deck always is prepared in the same order (A♠ 2♠ $\cdots$ K♣).

(a) What is the information entropy of the distribution of new decks?

(b) What is the information entropy of a distribution of completely randomized decks?

Now consider what it means to shuffle the cards. In an ideal *riffle shuffle*, the deck is split and divided into two equal halves of 26 cards each. One then chooses at random whether to take a card from either half, until one runs through all the cards and a new order is established (see figure).

Figure 1: The riffle shuffle.

(c) What is the increase in information entropy for a distribution of new decks that each have been shuffled once?

(d) Assuming each subsequent shuffle results in the same entropy increase (*i.e.* neglecting redundancies), how many shuffles are necessary in order to completely randomize a deck?

Solution :

(a) Since each new deck arrives in the same order, we have $p_1 = 1$ while $p_{2,\ldots,52!} = 0$, so $S = 0$.

(b) For completely randomized decks, $p_n = 1/\Omega$ with $n \in \{1, \ldots, \Omega\}$ and $\Omega = 52!$, the total number of possible configurations. Thus, $S_{\text{random}} = \log_2 52! = 225.581$.

(c) After one riffle shuffle, there are $\Omega = \binom{52}{26}$ possible configurations. If all such configurations were equally likely, we would have $(\Delta S)_{\text{riffle}} = \log_2 \binom{52}{26} = 48.817$. However, they are not all equally likely. For example, the probability that we drop the entire left-half deck and then the entire right half-deck is 2^{-26}. After the last card from the left half-deck is dropped, we have no more choices to make. On the other hand, the probability for the sequence LRLR$\cdots$ is 2^{-51}, because it is only after the 51^{st} card is dropped that we have no more choices. We can derive an exact expression for the entropy of the riffle shuffle in the following manner. Consider a deck of $N = 2K$ cards. The probability that we run out of choices after K cards is the probability of the first K cards dropped being all from one particular half-deck, which is $2 \cdot 2^{-K}$. Now let's ask what is the probability that we run out of choices after $(K+1)$ cards are dropped. If all the remaining $(K-1)$ cards are from the right half-deck, this means that we must have one of the R cards among the first K dropped. Note that this R card cannot be the $(K+1)^{\text{th}}$ card dropped, since then all of the first K cards are L, which we have already considered. Thus, there are $\binom{K}{1} = K$ such configurations, each with a probability 2^{-K-1}. Next, suppose we run out of choices after $(K+2)$ cards are dropped. If the remaining $(K-2)$ cards are R, this means we must have 2 of the R cards among the first $(K+1)$ dropped, which means $\binom{K+1}{2}$ possibilities. Note that the $(K+2)^{\text{th}}$ card must be L, since if it were R this would mean that the last $(K-1)$ cards are R, which we have already considered. Continuing in this manner, we conclude

$$\Omega_K = 2\sum_{n=0}^{K} \binom{K+n-1}{n} = \binom{2K}{K}$$

K	Ω_K	S_K	$\log_2\binom{2K}{K}$
2	6	2.500	2.585
12	2704156	20.132	20.367
26	4.96×10^{14}	46.274	48.817
100	9.05×10^{58}	188.730	195.851

Table 1: Riffle shuffle results.

and

$$S_K = -\sum_{a=1}^{\Omega_K} p_a \log_2 p_a = \sum_{n=0}^{K-1} \binom{K+n-1}{n} \cdot 2^{-(K+n)} \cdot (K+n) \quad .$$

The results are tabulated below in Table 1. For a deck of 52 cards, the actual entropy per riffle shuffle is $S_{26} = 46.274$.

(d) Ignoring redundancies, we require $k = S_{\text{random}}/(\Delta S)_{\text{riffle}} = 4.62$ shuffles if we assume all riffle outcomes are equally likely, and 4.88 if we use the exact result for the riffle entropy. Since there are no fractional shuffles, we round up to $k = 5$ in both cases. In fact, computer experiments show that the answer is $k = 9$. The reason we are so far off is that we have ignored redundancies, *i.e.* we have assumed that all the states produced by two consecutive riffle shuffles are distinct. They are not! For decks with asymptotically large numbers of cards $N \gg 1$, the number of riffle shuffles required is $k \simeq \frac{3}{2} \log_2 N$. See D. Bayer and P. Diaconis, *Annals of Applied Probability* **2**, 294 (1992).

(1.2) A six-sided die is loaded so that the probability to throw a three is twice that of throwing a two, and the probability of throwing a four is twice that of throwing a five.

(a) Find the distribution $\{p_n\}$ consistent with maximum entropy, given these constraints.

(b) Assuming the maximum entropy distribution, given two such identical dice, what is the probability to roll a total of seven if both are thrown simultaneously?

Solution :

(a) We have the following constraints:

$$\begin{aligned} X^0(\boldsymbol{p}) &= p_1 + p_2 + p_3 + p_4 + p_5 + p_6 - 1 = 0 \\ X^1(\boldsymbol{p}) &= p_3 - 2p_2 = 0 \\ X^2(\boldsymbol{p}) &= p_4 - 2p_5 = 0 \quad . \end{aligned}$$

We define

$$S^*(\boldsymbol{p}, \boldsymbol{\lambda}) \equiv -\sum_n p_n \ln p_n - \sum_{a=0}^{2} \lambda_a \, X^{(a)}(\boldsymbol{p}) \quad ,$$

and freely extremize over the probabilities $\{p_1, \ldots, p_6\}$ and the undetermined Lagrange multipliers $\{\lambda_0, \lambda_1, \lambda_2\}$. We obtain

$$\begin{aligned} \frac{\partial S^*}{\partial p_1} &= -1 - \ln p_1 - \lambda_0 & \qquad \frac{\partial S^*}{\partial p_4} &= -1 - \ln p_4 - \lambda_0 - \lambda_2 \\ \frac{\partial S^*}{\partial p_2} &= -1 - \ln p_2 - \lambda_0 + 2\lambda_1 & \qquad \frac{\partial S^*}{\partial p_5} &= -1 - \ln p_5 - \lambda_0 + 2\lambda_2 \\ \frac{\partial S^*}{\partial p_3} &= -1 - \ln p_3 - \lambda_0 - \lambda_1 & \qquad \frac{\partial S^*}{\partial p_6} &= -1 - \ln p_6 - \lambda_0 \quad . \end{aligned}$$

Extremizing with respect to the undetermined multipliers generates the three constraint equations. We therefore have

$$\begin{aligned} p_1 &= e^{-\lambda_0 - 1} & \qquad p_4 &= e^{-\lambda_0 - 1} \, e^{-\lambda_2} \\ p_2 &= e^{-\lambda_0 - 1} \, e^{2\lambda_1} & \qquad p_5 &= e^{-\lambda_0 - 1} \, e^{2\lambda_2} \\ p_3 &= e^{-\lambda_0 - 1} \, e^{-\lambda_1} & \qquad p_6 &= e^{-\lambda_0 - 1} \quad . \end{aligned}$$

We solve for $\{\lambda_0, \lambda_1, \lambda_2\}$ by imposing the three constraints. Let $x \equiv p_1 = p_6 = e^{-\lambda_0 - 1}$. Then $p_2 = x\, e^{2\lambda_1}$, $p_3 = x\, e^{-\lambda_1}$, $p_4 = x\, e^{-\lambda_2}$, and $p_5 = x\, e^{2\lambda_2}$. We then have

$$\begin{aligned} p_3 = 2p_2 \quad &\Rightarrow \quad e^{-3\lambda_1} = 2 \\ p_4 = 2p_5 \quad &\Rightarrow \quad e^{-3\lambda_2} = 2 \quad . \end{aligned}$$

We may now solve for x:

$$\sum_{n=1}^{6} p_n = \left(2 + 2^{1/3} + 2^{4/3}\right) x = 1 \quad \Rightarrow \quad x = \frac{1}{2 + 3 \cdot 2^{1/3}} \quad .$$

We now have all the probabilities:

$$\begin{aligned} p_1 &= x = 0.1730 & \qquad p_4 &= 2^{1/3} x = 0.2180 \\ p_2 &= 2^{-2/3} x = 0.1090 & \qquad p_5 &= 2^{-2/3} x = 0.1090 \\ p_3 &= 2^{1/3} x = 0.2180 & \qquad p_6 &= x = 0.1730 \quad . \end{aligned}$$

(b) The probability to roll a seven with two of these dice is

$$\begin{aligned} P(7) &= 2\, p_1\, p_6 + 2\, p_2\, p_5 + 2\, p_3\, p_4 \\ &= 2\left(1 + 2^{-4/3} + 2^{2/3}\right) x^2 = 0.1787 \quad . \end{aligned}$$

(1.3) Consider the contraption in Fig. 2. At each of k steps, a particle can fork to either the left ($n_j = 1$) or to the right ($n_j = 0$). The final location is then a k-digit binary number.

(a) Assume the probability for moving to the left is p and the probability for moving to the right is $q \equiv 1 - p$ at each fork, independent of what happens at any of the other forks. *I.e.* all the forks are uncorrelated. Compute $\langle X_k \rangle$. *Hint: X_k can be represented as a k-digit binary number, i.e.* $X_k = n_{k-1}n_{k-2}\cdots n_1 n_0 = \sum_{j=0}^{k-1} 2^j n_j$.

(b) Compute $\langle X_k^2 \rangle$ and the variance $\langle X_k^2 \rangle - \langle X_k \rangle^2$.

(c) X_k may be written as the sum of k random numbers. Does X_k satisfy the central limit theorem as $k \to \infty$? Why or why not?

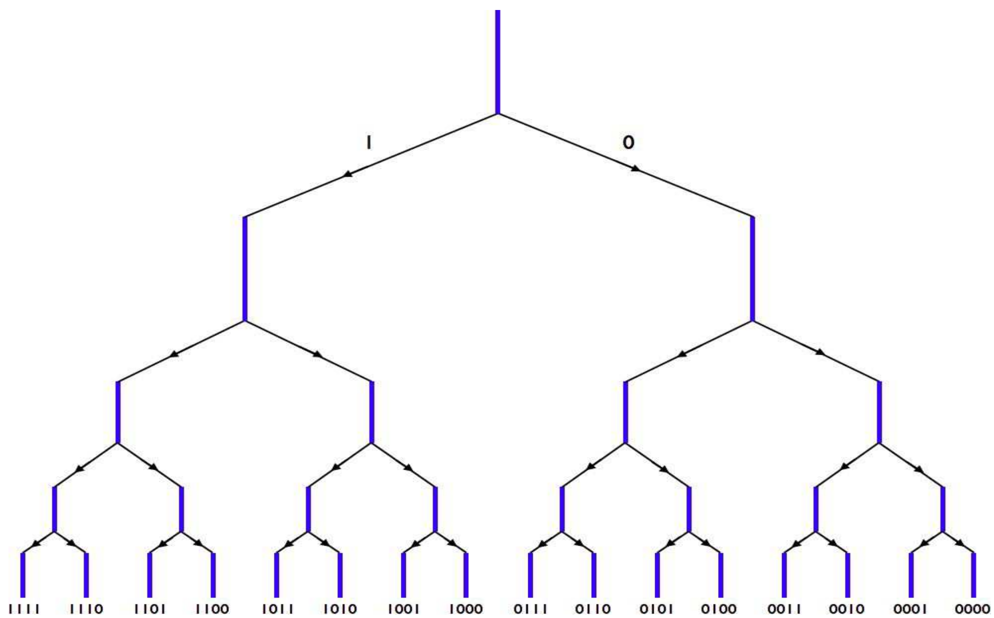


Figure 2: Generator for a k-digit random binary number ($k = 4$ shown).

Solution :

(a) The position after k forks can be written as a k-digit binary number: $n_{k-1}n_{k-2}\cdots n_1 n_0$. Thus,

$$X_k = \sum_{j=0}^{k-1} 2^j\, n_j \quad ,$$

where $n_j = 0$ or 1 according to $P_n = p\,\delta_{n,1} + q\,\delta_{n,0}$. Now it is clear that $\langle n_j \rangle = p$, and therefore

$$\langle X_k \rangle = p \sum_{j=0}^{k-1} 2^j = p \cdot \big(2^k - 1\big) \quad .$$

(b) The variance in X_k is

$$\begin{aligned} \mathrm{Var}(X_k) = \langle X_k^2 \rangle - \langle X_k \rangle^2 &= \sum_{j=0}^{k-1} \sum_{j'=0}^{k-1} 2^{j+j'} \Big(\langle n_j n_{j'} \rangle - \langle n_j \rangle \langle n_{j'} \rangle \Big) \\ &= p(1-p) \sum_{j=0}^{k-1} 4^j = p(1-p) \cdot \tfrac{1}{3}\big(4^k - 1\big) \quad , \end{aligned}$$

since $\langle n_j n_{j'} \rangle - \langle n_j \rangle \langle n_{j'} \rangle = p(1-p)\, \delta_{jj'}$.

(c) Clearly the distribution of X_k does not obey the CLT, since $\langle X_k \rangle$ scales exponentially with k. Also note

$$\lim_{k\to\infty} \frac{\sqrt{\mathrm{Var}(X_k)}}{\langle X_k \rangle} = \sqrt{\frac{1-p}{3p}} \quad ,$$

which is a constant. For distributions obeying the CLT, the ratio of the rms fluctuations to the mean scales as the inverse square root of the number of trials. The reason that this distribution does not obey the CLT is that the variance of the individual terms is increasing with j.

(1.4) The binomial distribution,

$$B_N(n,p) = \binom{N}{n} p^n \, (1-p)^{N-n} \quad ,$$

tells us the probability for n successes in N trials if the individual trial success probability is p. The average number of successes is $\nu = \sum_{n=0}^{N} n \, B_N(n,p) = Np$. Consider the limit $N \to \infty$ with ν finite.

(a) Show that the probability of n successes becomes a function of n and ν alone. That is, evaluate

$$P_\nu(n) = \lim_{N\to\infty} B_N(n, \nu/N) \quad .$$

This is the *Poisson distribution*.

(b) Show that the moments of the Poisson distribution are given by

$$\langle n^k \rangle = e^{-\nu} \left(\nu \frac{\partial}{\partial \nu} \right)^k e^{\nu} \quad .$$

(c) Evaluate the mean and variance of the Poisson distribution.

The Poisson distribution is also known as the *law of rare events* since $p = \nu/N \to 0$ in the $N \to \infty$ limit. See http://en.wikipedia.org/wiki/Poisson_distribution#Occurrence for some amusing applications of the Poisson distribution.

Solution :

(a) We have

$$P_\nu(n) = \lim_{N\to\infty} \frac{N!}{n!\,(N-n)!} \left(\frac{\nu}{N}\right)^n \left(1 - \frac{\nu}{N}\right)^{N-n} \quad .$$

Note that

$$(N-n)! \simeq (N-n)^{N-n} \, e^{n-N} = N^{N-n} \left(1 - \frac{n}{N}\right)^N e^{n-N} \to N^{N-n} \, e^N \quad ,$$

where we have used the result $\lim_{N\to\infty} \left(1 + \frac{x}{N}\right)^N = e^x$. Thus, we find

$$P_\nu(n) = \frac{1}{n!} \, \nu^n \, e^{-\nu} \quad ,$$

the Poisson distribution. Note that $\sum_{n=0}^{\infty} P_n(\nu) = 1$ for any ν.

(b) We have

$$\begin{aligned} \langle n^k \rangle &= \sum_{n=0}^{\infty} P_\nu(n) \, n^k = \sum_{n=0}^{\infty} \frac{1}{n!} \, n^k \nu^n \, e^{-\nu} \\ &= e^{-\nu} \left(\nu \frac{d}{d\nu} \right)^k \sum_{n=0}^{\infty} \frac{\nu^n}{n!} = e^{-\nu} \left(\nu \frac{\partial}{\partial \nu} \right)^k e^{\nu} \quad . \end{aligned}$$

(c) Using the result from (b), we have $\langle n \rangle = \nu$ and $\langle n^2 \rangle = \nu + \nu^2$, hence $\mathrm{Var}(n) = \nu$.

(1.5) You should be familiar with Stirling's approximation,

$$\ln K! \sim K \ln K - K + \tfrac{1}{2}\ln(2\pi K) + \mathcal{O}\big(K^{-1}\big) \quad ,$$

for large K. In this exercise, you will derive this expansion.

(a) Start by writing

$$K! = \int_0^\infty dx\, x^K\, e^{-x} \quad ,$$

and define $x \equiv K(t+1)$ so that $K! = K^{K+1}\, e^{-K}\, F(K)$, where

$$F(K) = \int_{-1}^\infty dt\, e^{K f(t)} \quad .$$

Find the function $f(t)$.

(b) Expand $f(t) = \sum_{n=0}^\infty f_n\, t^n$ in a Taylor series and find a general formula for the expansion coefficients f_n. In particular, show that $f_0 = f_1 = 0$ and that $f_2 = -\frac{1}{2}$.

(c) If one ignores all the terms but the lowest order (quadratic) in the expansion of $f(t)$, show that

$$\int_{-1}^\infty dt\, e^{-Kt^2/2} = \sqrt{\frac{2\pi}{K}} - R(K) \quad ,$$

and show that the remainder $R(K) > 0$ is bounded from above by a function which decreases faster than any polynomial in $1/K$.

(d) *For the brave only!* – Find the $\mathcal{O}\big(K^{-1}\big)$ term in the expansion for $\ln K!$.

Solution :

(a) Setting $x = K(t+1)$, we have

$$K! = K^{K+1}\, e^{-K} \int_{-1}^\infty dt\, (t+1)^K\, e^{-t} \quad ,$$

hence $f(t) = \ln(t+1) - t$.

(b) The Taylor expansion of $f(t)$ is

$$f(t) = -\tfrac{1}{2}t^2 + \tfrac{1}{3}t^3 - \tfrac{1}{4}t^4 + \ldots \quad .$$

(c) Retaining only the leading term in the Taylor expansion of $f(t)$, we have

$$\begin{aligned} F(K) &\simeq \int_{-1}^\infty dt\, e^{-Kt^2/2} \\ &= \sqrt{\frac{2\pi}{K}} - \int_1^\infty dt\, e^{-Kt^2/2} \quad . \end{aligned}$$

Writing $t \equiv s + 1$, the remainder is found to be

$$R(K) = e^{-K/2} \int\limits_0^\infty ds\, e^{-Ks^2/2}\, e^{-Ks} < \sqrt{\frac{\pi}{2K}}\, e^{-K/2} \quad ,$$

which decreases exponentially with K, faster than any power.

(d) We have

$$\begin{aligned} F(K) &= \int\limits_{-1}^\infty dt\, e^{-\frac{1}{2}Kt^2} e^{\frac{1}{3}Kt^3 - \frac{1}{4}Kt^4 + \ldots} \\ &= \int\limits_{-1}^\infty dt\, e^{-\frac{1}{2}Kt^2} \Big\{ 1 + \tfrac{1}{3}Kt^3 - \tfrac{1}{4}Kt^4 + \tfrac{1}{18}K^2t^6 + \ldots \Big\} \\ &= \sqrt{\frac{2\pi}{K}} \cdot \Big\{ 1 - \tfrac{3}{4}K^{-1} + \tfrac{5}{6}K^{-1} + \mathcal{O}\big(K^{-2}\big) \Big\} \end{aligned}$$

Thus,

$$\ln K! = K \ln K - K + \tfrac{1}{2}\ln K + \tfrac{1}{2}\ln(2\pi) + \tfrac{1}{12}K^{-1} + \mathcal{O}\big(K^{-2}\big) \quad .$$

(1.6) The probability density for a random variable x is given by the Lorentzian,

$$P(x) = \frac{\gamma}{\pi} \cdot \frac{1}{x^2 + \gamma^2} \quad .$$

Consider the sum $X_N = \sum_{i=1}^{N} x_i$, where each x_i is independently distributed according to $P(x_i)$. Find the probability $\mathit{\Pi}_N(Y)$ that $|X_N| < Y$, where $Y > 0$ is arbitrary.

Solution :

The distribution of a sum of identically distributed random variables, $X = \sum_{i=1}^{N} x_i$, is given by

$$P_N(X) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} \left[\hat{P}(k)\right]^N e^{ikX} \quad ,$$

where $\hat{P}(k)$ is the Fourier transform of the probability distribution $P(x_i)$ for each of the x_i. The Fourier transform of a Lorentzian is an exponential:

$$\int_{-\infty}^{\infty} dx \, P(x) \, e^{-ikx} = e^{-\gamma|k|} \quad .$$

Thus,

$$P_N(X) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} \, e^{-N\gamma|k|} \, e^{ikX} = \frac{N\gamma}{\pi} \cdot \frac{1}{X^2 + N^2\gamma^2} \quad .$$

The probability for X to lie in the interval $X \in [-Y, Y]$, where $Y > 0$, is

$$\mathit{\Pi}_N(Y) = \int_{-Y}^{Y} dX \, P_N(X) = \frac{2}{\pi} \tan^{-1}\left(\frac{Y}{N\gamma}\right) \quad .$$

The integral is easily performed with the substitution $X = N\gamma \tan\theta$. Note that $\mathit{\Pi}_N(0) = 0$ and $\mathit{\Pi}_N(\infty) = 1$.

(1.7) Let $P(x) = (2\pi\sigma^2)^{-1/2}\, e^{-(x-\mu)^2/2\sigma^2}$. Compute the following integrals:

(a) $I = \int\limits_{-\infty}^{\infty} dx\, P(x)\, x^3$.

(b) $I = \int\limits_{-\infty}^{\infty} dx\, P(x)\, \cos(Qx)$.

(c) $I = \int\limits_{-\infty}^{\infty} dx \int\limits_{-\infty}^{\infty} dy\, P(x)\, P(y)\, e^{xy}$. You may set $\mu = 0$ to make this somewhat simpler. Under what conditions does this expression converge?

Solution :

(a) Write

$$x^3 = (x - \mu + \mu)^3 = (x-\mu)^3 + 3(x-\mu)^2\mu + 3(x-\mu)\mu^2 + \mu^3 \quad ,$$

so that

$$\langle x^3 \rangle = \frac{1}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{\infty} dt\, e^{-t^2/2\sigma^2} \Big\{ t^3 + 3t^2\mu + 3t\mu^2 + \mu^3 \Big\} \quad .$$

Since $\exp(-t^2/2\sigma^2)$ is an even function of t, odd powers of t integrate to zero. We have $\langle t^2 \rangle = \sigma^2$, so

$$\langle x^3 \rangle = \mu^3 + 3\mu\sigma^2 \quad .$$

A nice trick for evaluating $\langle t^{2k} \rangle$:

$$\langle t^{2k} \rangle = \frac{\int\limits_{-\infty}^{\infty} dt\, e^{-\lambda t^2}\, t^{2k}}{\int\limits_{-\infty}^{\infty} dt\, e^{-\lambda t^2}} = \frac{(-1)^k \frac{d^k}{d\lambda^k} \int\limits_{-\infty}^{\infty} dt\, e^{-\lambda t^2}}{\int\limits_{-\infty}^{\infty} dt\, e^{-\lambda t^2}} = \frac{(-1)^k}{\sqrt{\lambda}} \frac{d^k \sqrt{\lambda}}{d\lambda^k} \bigg|_{\lambda = 1/2\sigma^2}$$

$$= \tfrac{1}{2} \cdot \tfrac{3}{2} \cdots \tfrac{(2k-1)}{2}\, \lambda^{-k} \Big|_{\lambda=1/2\sigma^2} = \frac{(2k)!}{2^k\, k!}\, \sigma^{2k} \quad .$$

(b) We have

$$\langle \cos(Qx) \rangle = \mathrm{Re}\, \langle e^{iQx} \rangle = \mathrm{Re} \left[\frac{e^{iQ\mu}}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{\infty} dt\, e^{-t^2/2\sigma^2}\, e^{iQt} \right]$$

$$= \mathrm{Re} \left[e^{iQ\mu}\, e^{-Q^2\sigma^2/2} \right] = \cos(Q\mu)\, e^{-Q^2\sigma^2/2} \quad .$$

Here we have used the result

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{\infty} dt\, e^{-\alpha t^2 - \beta t} = \sqrt{\frac{\pi}{\alpha}}\, e^{\beta^2/4\alpha}$$

with $\alpha = 1/2\sigma^2$ and $\beta = -iQ$. Another way to do it is to use the general result derive above in part (a) for $\langle t^{2k} \rangle$ and do the sum:

$$\langle \cos(Qx) \rangle = \mathrm{Re}\, \langle e^{iQx} \rangle = \mathrm{Re} \left[\frac{e^{iQ\mu}}{\sqrt{2\pi\sigma^2}} \int\limits_{-\infty}^{\infty} dt\, e^{-t^2/2\sigma^2}\, e^{iQt} \right]$$

$$= \cos(Q\mu) \sum_{k=0}^{\infty} \frac{(-Q^2)^k}{(2k)!} \langle t^{2k} \rangle = \cos(Q\mu) \sum_{k=0}^{\infty} \frac{1}{k!} \left(-\tfrac{1}{2} Q^2 \sigma^2 \right)^k$$

$$= \cos(Q\mu)\, e^{-Q^2\sigma^2/2} \quad .$$

(c) We have

$$I = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \, P(x) \, P(y) \, e^{\kappa^2 xy} = \frac{e^{-\mu^2/2\sigma^2}}{2\pi\sigma^2} \int d^2x \, e^{-\frac{1}{2} A_{ij} \, x_i \, x_j} \, e^{b_i \, x_i} \quad ,$$

where $\boldsymbol{x} = (x, y)$,

$$A = \begin{pmatrix} \sigma^2 & -\kappa^2 \\ -\kappa^2 & \sigma^2 \end{pmatrix} \quad , \quad \boldsymbol{b} = \begin{pmatrix} \mu/\sigma^2 \\ \mu/\sigma^2 \end{pmatrix} \quad .$$

Using the general formula for the Gaussian integral,

$$\int d^n x \, e^{-\frac{1}{2} A_{ij} \, x_i \, x_j} \, e^{b_i \, x_i} = \frac{(2\pi)^{n/2}}{\sqrt{\det(A)}} \, \exp\left(\tfrac{1}{2} A^{-1}_{ij} \, b_i \, b_j\right) \quad ,$$

we obtain

$$I = \frac{1}{\sqrt{1 - \kappa^4 \sigma^4}} \, \exp\left(\frac{\mu^2 \kappa^2}{1 - \kappa^2 \sigma^2}\right) \quad .$$

Convergence requires $\kappa^2 \sigma^2 < 1$.

(1.8) Consider a D-dimensional *random walk* on a hypercubic lattice. The position of a particle after N steps is

$$\boldsymbol{R}_N = \sum_{j=1}^{N} \hat{\boldsymbol{n}}_j \quad ,$$

where $\hat{\boldsymbol{n}}_j$ can take on one of $2D$ possible values: $\hat{\boldsymbol{n}}_j \in \{\pm\hat{\mathbf{e}}_1, \ldots, \pm\hat{\mathbf{e}}_D\}$, where $\hat{\mathbf{e}}_\mu$ is the unit vector along the positive x_μ axis. Each of these possible values occurs with probability $1/2D$, and each step is statistically independent from all other steps.

(a) Consider the *generating function* $S_N(\boldsymbol{k}) = \langle e^{i\boldsymbol{k}\cdot\boldsymbol{R}_N}\rangle$. Show that

$$\langle R_N^{\alpha_1} \cdots R_N^{\alpha_J}\rangle = \frac{1}{i}\frac{\partial}{\partial k_{\alpha_1}} \cdots \frac{1}{i}\frac{\partial}{\partial k_{\alpha_J}}\bigg|_{\boldsymbol{k}=0} S_N(\boldsymbol{k}) \quad .$$

For example, $\langle R_N^\alpha R_N^\beta\rangle = -\big(\partial^2 S_N(\boldsymbol{k})/\partial k_\alpha \partial k_\beta\big)_{\boldsymbol{k}=0}$.

(b) Evaluate $S_N(\boldsymbol{k})$ for the case $D=3$ and compute the quantities $\langle X_N^4\rangle$ and $\langle X_N^2\, Y_N^2\rangle$.

Solution :

(a) The result follows immediately from

$$\frac{1}{i}\frac{\partial}{\partial k_\alpha}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} = R_\alpha\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}}$$
$$\frac{1}{i}\frac{\partial}{\partial k_\alpha}\,\frac{1}{i}\frac{\partial}{\partial k_\beta}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} = R_\alpha\, R_\beta\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \quad ,$$

et cetera. Keep differentiating with respect to the various components of $\boldsymbol{k}$.

(b) For $D=3$, there are six possibilities for $\hat{\boldsymbol{n}}_j$: $\pm\hat{\boldsymbol{x}}$, $\pm\hat{\boldsymbol{y}}$, and $\pm\hat{\boldsymbol{z}}$. Each occurs with a probability $\frac{1}{6}$, independent of all the other $\hat{\boldsymbol{n}}_{j'}$ with $j' \neq j$. Thus,

$$\begin{aligned} S_N(\boldsymbol{k}) &= \prod_{j=1}^{N} \langle e^{i\boldsymbol{k}\cdot\hat{\boldsymbol{n}}_j}\rangle = \left[\frac{1}{6}\Big(e^{ik_x} + e^{-ik_x} + e^{ik_y} + e^{-ik_y} + e^{ik_z} + e^{-ik_z}\Big)\right]^N \\ &= \left(\frac{\cos k_x + \cos k_y + \cos k_z}{3}\right)^N \quad . \end{aligned}$$

We have

$$\begin{aligned} \langle X_N^4\rangle &= \frac{\partial^4 S(\boldsymbol{k})}{\partial k_x^4}\bigg|_{\boldsymbol{k}=0} = \frac{\partial^4}{\partial k_x^4}\bigg|_{k_x=0} \big(1 - \tfrac{1}{6}\,k_x^2 + \tfrac{1}{72}\,k_x^4 + \ldots\big)^N \\ &= \frac{\partial^4}{\partial k_x^4}\bigg|_{k_x=0} \Big[1 + N\big(-\tfrac{1}{6}\,k_x^2 + \tfrac{1}{72}\,k_x^4 + \ldots\big) + \tfrac{1}{2}N(N-1)\big(-\tfrac{1}{6}\,k_x^2 + \tfrac{1}{72}\,k_x^4 + \ldots\big)^2 + \ldots\Big] \\ &= \frac{\partial^4}{\partial k_x^4}\bigg|_{k_x=0} \Big[1 - \tfrac{1}{6}Nk_x^2 + \tfrac{1}{72}N^2 k_x^4 + \ldots\Big] = \tfrac{1}{3}N^2 \quad . \end{aligned}$$

Similarly, we have

$$
\begin{aligned}
\langle X_N^2\, Y_N^2 \rangle = \frac{\partial^4 S(\boldsymbol{k})}{\partial k_x^2\, \partial k_y^2}\bigg|_{\boldsymbol{k}=0} &= \frac{\partial^4}{\partial k_x^2\, \partial k_y^2}\bigg|_{k_x=0} \big(1 - \tfrac{1}{6}\,(k_x^2 + k_y^2) + \tfrac{1}{72}\,(k_x^4 + k_y^4) + \ldots\big)^N \\
&= \frac{\partial^4}{\partial k_x^2\, \partial k_y^2}\bigg|_{k_x=k_y=0} \Big[1 + N\big(-\tfrac{1}{6}\,(k_x^2 + k_y^2) + \tfrac{1}{72}\,(k_x^4 + k_y^4) + \ldots\big) + \tfrac{1}{2}N(N-1)\big(-\tfrac{1}{6}\,(k_x^2 + k_y^2) + \ldots\big)^2 + \ldots\Big] \\
&= \frac{\partial^4}{\partial k_x^2\, \partial k_y^2}\bigg|_{k_x=k_y=0} \Big[1 - \tfrac{1}{6}N(k_x^2 + k_y^2) + \tfrac{1}{72}N^2(k_x^4 + k + y^4) + \tfrac{1}{36}\,k_x^2\,k_y^2 + \ldots\Big] = \tfrac{1}{9}N(N-1) \quad .
\end{aligned}
$$

(1.9) A rare disease is known to occur in $f = 0.02\%$ of the general population. Doctors have designed a test for the disease with $\nu = 99.90\%$ sensitivity and $\rho = 99.95\%$ specificity.

(a) What is the probability that someone who tests positive for the disease is actually sick?

(b) Suppose the test is administered twice, and the results of the two tests are independent. If a random individual tests positive both times, what are the chances he or she actually has the disease?

(c) For a binary partition of events, find an expression for $P(X|A \cap B)$ in terms of $P(A|X)$, $P(B|X)$, $P(A|\neg X)$, $P(B|\neg X)$, and the priors $P(X)$ and $P(\neg X) = 1 - P(X)$. You should assume A and B are independent, so $P(A \cap B|X) = P(A|X) \cdot P(B|X)$.

Solution :

(a) Let X indicate that a person is infected, and A indicate that a person has tested positive. We then have $\nu = P(A|X) = 0.9990$ is the sensitivity and $\rho = P(\neg A|\neg X) = 0.9995$ is the specificity. From Bayes' theorem, we have

$$P(X|A) = \frac{P(A|X) \cdot P(X)}{P(A|X) \cdot P(X) + P(A|\neg X) \cdot P(\neg X)} = \frac{\nu f}{\nu f + (1-\rho)(1-f)} \quad ,$$

where $P(A|\neg X) = 1 - P(\neg A|\neg X) = 1 - \rho$ and $P(X) = f$ is the fraction of infected individuals in the general population. With $f = 0.0002$, we find $P(X|A) = 0.2856$.

(b) We now need

$$P(X|A^2) = \frac{P(A^2|X) \cdot P(X)}{P(A^2|X) \cdot P(X) + P(A^2|\neg X) \cdot P(\neg X)} = \frac{\nu^2 f}{\nu^2 f + (1-\rho)^2(1-f)} \quad ,$$

where A^2 indicates two successive, independent tests. We find $P(X|A^2) = 0.9987$.

(c) Assuming A and B are independent, we have

$$\begin{aligned} P(X|A \cap B) &= \frac{P(A \cap B|X) \cdot P(X)}{P(A \cap B|X) \cdot P(X) + P(A \cap B|\neg X) \cdot P(\neg X)} \\ &= \frac{P(A|X) \cdot P(B|X) \cdot P(X)}{P(A|X) \cdot P(B|X) \cdot P(X) \ + P(A|\neg X) \cdot P(B|\neg X) \cdot P(\neg X)} \quad . \end{aligned}$$

This is exactly the formula used in part (b).

(1.10) Let $p(x) = \mathsf{Pr}[X = x]$ where X is a discrete random variable and both X and x are taken from an 'alphabet' $\mathcal{X}$. Let $p(x,y)$ be a normalized joint probability distribution on two random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$. The entropy of the joint distribution is $S(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y)$. The conditional probability $p(y|x)$ for y given x is defined as $p(y|x) = p(x,y)/p(x)$, where $p(x) = \sum_y p(x,y)$.

(a) Show that the conditional entropy $S(Y|X) = -\sum_{x,y} p(x,y) \log p(y|x)$ satisfies

$$S(Y|X) = S(X,Y) - S(X) \quad .$$

Thus, conditioning reduces the entropy, and the entropy of a pair of random variables is the sum of the entropy of one plus the conditional entropy of the other.

(b) The mutual information $I(X,Y)$ is

$$I(X,Y) = \sum_{x,y} p(x,y) \log\left(\frac{p(x,y)}{p(x)\,p(y)}\right) \quad .$$

Show that

$$I(X,Y) = S(X) + S(Y) - S(X,Y) \quad .$$

(c) Show that $S(Y|X) \le S(Y)$ and that the equality holds only when X and Y are independently distributed.

Solution :

(a) We have

$$\begin{aligned} S(Y|X) &= -\sum_{x,y} p(x,y) \log\left(\frac{p(x,y)}{p(x)}\right) \\ &= -\sum_{x,y} p(x,y) \log p(x,y) + \sum_{x,y} p(x,y) \log p(x) \\ &= S(X,Y) - S(X) \quad , \end{aligned}$$

since $\sum_y p(x,y) = p(x)$.

(b) Clearly

$$\begin{aligned} I(X,Y) &= \sum_{x,y} p(x,y) \Big(\log p(x,y) - \log p(x) - \log p(y) \Big) \\ &= \sum_{x,y} p(x,y) \log p(x,y) - \sum_{x} p(x) \log p(x) - \sum_{y} p(y) \log p(y) \\ &= S(X) + S(Y) - S(X,Y) \quad . \end{aligned}$$

(c) Let's introduce some standard notation[1]. Denote the average of a function of a random variable as

$$\mathbb{E} f(X) = \sum_{x \in \mathcal{X}} p(x) f(x) \quad .$$

Equivalently, we could use the familiar angular bracket notation for averages and write $\mathbb{E} f(X)$ as $\langle f(X) \rangle$. For averages over several random variables, we have

$$\mathbb{E} f(X,Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x,y)\, f(x,y) \quad ,$$

[1] See *e.g.* T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd edition (Wiley, 2006).

et cetera. Now here's a useful fact. If $f(x)$ is a convex function, then

$$\mathbb{E} f(X) \geq f(\mathbb{E} X) \quad .$$

For continuous functions, $f(x)$ is convex if $f''(x) \geq 0$ everywhere[2]. If $f(x)$ is convex on some interval $[a, b]$, then for $x_{1,2} \in [a, b]$ we must have

$$f\big(\lambda\, x_1 + (1-\lambda)\, x_2\big) \leq \lambda\, f(x_1) + (1-\lambda)\, f(x_2) \quad .$$

This is easily generalized to

$$f\Big(\sum_n p_n\, x_n\Big) \leq \sum_n p_n\, f(x_n) \quad ,$$

where $\sum_n p_n = 1$, which proves our claim - a result known as *Jensen's theorem*.

Now $\log x$ is concave, hence the function $f(x) = -\log x$ is convex, and therefore

$$\begin{aligned} I(X,Y) &= -\sum_{x,y} p(x,y) \log\bigg(\frac{p(x)\, p(y)}{p(x,y)}\bigg) = -\mathbb{E} \log\bigg(\frac{p(X)\, p(Y)}{p(X,Y)}\bigg) \\ &\geq -\log \mathbb{E}\bigg(\frac{p(X)\, p(Y)}{p(X,Y)}\bigg) = -\log\bigg(\sum_{x,y} p(x,y) \cdot \frac{p(x)\, p(y)}{p(x,y)}\bigg) \\ &= -\log\bigg(\sum_x p(x) \cdot \sum_y p(y)\bigg) = -\log(1) = 0 \quad . \end{aligned}$$

So $I(X,Y) \geq 0$. Clearly $I(X,Y) = 0$ when X and Y are independently distributed, *i.e.* when $p(x,y) = p(x)\, p(y)$. Using the results from part (b), we then have

$$\begin{aligned} I(X,Y) &= S(X) + S(Y) - S(X,Y) \\ &= S(Y) - S(Y|X) \geq 0 \quad . \end{aligned}$$

[2]A *concave* function $g(x)$ is one for which $f(x) = -g(x)$ is convex.

(1.11) The n^{th} moment of the normalized Gaussian distribution $P(x) = (2\pi)^{-1/2} \exp\big(-\frac{1}{2}x^2\big)$ is defined by

$$\langle x^n \rangle = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx\; x^n \, \exp\big(-\tfrac{1}{2}x^2\big)$$

Clearly $\langle x^n \rangle = 0$ if n is a nonnegative odd integer. Next consider the *generating function*

$$Z(j) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \, \exp\big(-\tfrac{1}{2}x^2\big)\exp(jx) = \exp\big(\tfrac{1}{2}j^2\big) \quad .$$

(a) Show that

$$\langle x^n \rangle = \frac{d^n Z}{dj^n}\bigg|_{j=0}$$

and provide an explicit result for $\langle x^{2k} \rangle$ where $k \in \mathbb{N}$.

(b) Now consider the following integral:

$$F(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \, \exp\bigg(-\frac{1}{2}\,x^2 - \frac{\lambda}{4!}\,x^4\bigg) \quad .$$

The integral has no known analytic form[3], but we may express the result as a power series in the parameter λ by Taylor expanding $\exp\big(-\frac{\lambda}{4!}\,x^4\big)$ and then using the result of part (a) for the moments $\langle x^{4k} \rangle$. Find the coefficients in the perturbation expansion,

$$F(\lambda) = \sum_{k=0}^{\infty} C_k \, \lambda^k \quad .$$

(c) Define the *remainder after N terms* as

$$R_N(\lambda) = F(\lambda) - \sum_{k=0}^{N} C_k \, \lambda^k \quad .$$

Compute $R_N(\lambda)$ by evaluating numerically the integral for $F(\lambda)$ (using `Mathematica` or some other numerical package) and subtracting the finite sum. Then define the ratio $S_N(\lambda) = R_N(\lambda)/F(\lambda)$, which is the relative error from the N term approximation and plot the absolute relative error $\big|S_N(\lambda)\big|$ *versus* N for several values of λ.(I suggest you plot the error on a log scale.) What do you find?? Try a few values of λ including $\lambda = 0.01$, $\lambda = 0.05$, $\lambda = 0.2$, $\lambda = 0.5$, $\lambda = 1$, $\lambda = 2$.

(d) Repeat the calculation for the integral

$$G(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \, \exp\bigg(-\frac{1}{2}\,x^2 - \frac{\lambda}{6!}\,x^6\bigg) \quad .$$

(e) Reflect meaningfully on the consequences for weakly and strongly coupled quantum field theories.

[3] In fact, it does. According to `Mathematica`, $F(\lambda) = \sqrt{\frac{2u}{\pi}}\,\exp(u)\,K_{1/4}(u)$, where $u = 3/4\lambda$ and $K_\nu(z)$ is the modified Bessel function. I am grateful to Prof. John McGreevy for pointing this out.

Solution :

(a) Clearly

$$\frac{d^n}{dj^n}\bigg|_{j=0} e^{jx} = x^n \quad ,$$

so $\langle\, x^n \,\rangle = \big(d^n Z/dj^n\big)_{j=0}$. With $Z(j) = \exp\big(\frac{1}{2}j^2\big)$, only the $k^{\rm th}$ order term in j^2 in the Taylor series for $Z(j)$ contributes, and we obtain

$$\langle\, x^{2k} \,\rangle = \frac{d^{2k}}{dj^{2k}}\left(\frac{j^{2k}}{2^k\, k!}\right) = \frac{(2k)!}{2^k\, k!} \quad .$$

(b) We have

$$F(\lambda) = \sum_{n=0}^{\infty} \frac{1}{n!}\left(-\frac{\lambda}{4!}\right)^n \langle\, x^{4n} \,\rangle = \sum_{n=0}^{\infty} \frac{(4n)!}{4^n\,(4!)^n\, n!\,(2n)!}\,(-\lambda)^n \quad .$$

This series is *asymptotic*. It has the properties

$$\lim_{\lambda\to 0} \frac{R_N(\lambda)}{\lambda^N} = 0 \quad (\text{fixed } N) \qquad , \qquad \lim_{N\to\infty} \frac{R_N(\lambda)}{\lambda^N} = \infty \quad (\text{fixed } \lambda) \quad ,$$

where $R_N(\lambda)$ is the remainder after N terms, defined in part (c). The radius of convergence is zero. To see this, note that if we reverse the sign of λ, then the integrand of $F(\lambda)$ diverges badly as $x \to \pm\infty$. So $F(\lambda)$ is infinite for $\lambda < 0$, which means that there is no disk of any finite radius of convergence which encloses the point $\lambda = 0$. Note that by Stirling's rule,

$$(-1)^n\, C_n \equiv \frac{(4n)!}{4^n\,(4!)^n\, n!\,(2n)!} \sim n^n \cdot \big(\tfrac{2}{3}\big)^n\, e^{-n} \cdot (\pi n)^{-1/2} \quad ,$$

and we conclude that the magnitude of the summand reaches a minimum value when $n = n^*(\lambda)$, with

$$n^*(\lambda) \approx \frac{3}{2\lambda}$$

for small values of λ. For large n, the magnitude of the coefficient C_n grows as $|C_n| \sim e^{n\ln n + \mathcal{O}(n)}$, which dominates the λ^n term, no matter how small λ is.

(c) Results are plotted in fig. 3.

It is worth pointing out that the series for $F(\lambda)$ and for $\ln F(\lambda)$ have diagrammatic interpretations. For a Gaussian integral, one has

$$\langle\, x^{2k} \,\rangle = \langle\, x^2 \,\rangle^k \cdot A_{2k}$$

where A_{2k} is the *number of contractions*. For a proof, see §1.4.3 of the notes. For our integral, $\langle\, x^2 \,\rangle = 1$. The number of contractions A_{2k} is computed in the following way. For each of the $2k$ powers of x, we assign an index running from 1 to $2k$. The indices are *contracted*, *i.e.* paired, with each other. How many pairings are there? Suppose we start with any from among the $2k$ indices. Then there are $(2k-1)$ choices for its mate. We then choose another index arbitrarily. There are now $(2k-3)$ choices for *its* mate. Carrying this out to its completion, we find that the number of contractions is

$$A_{2k} = (2k-1)(2k-3)\cdots 3\cdot 1 = \frac{(2k)!}{2^k\, k!} \quad ,$$

exactly as we found in part (a). Now consider the integral $F(\lambda)$. If we expand the quartic term in a power series, then each power of λ brings an additional four powers of x. It is therefore convenient to represent each such quartet with the symbol $\times$. At order N of the series expansion, we have N $\times$'s and $4N$ indices to contract. Each full contraction of the indices may be represented as a labeled diagram, which is in general composed of several disjoint connected subdiagrams. Let us label these subdiagrams, which we will call clusters, by an index γ. Now

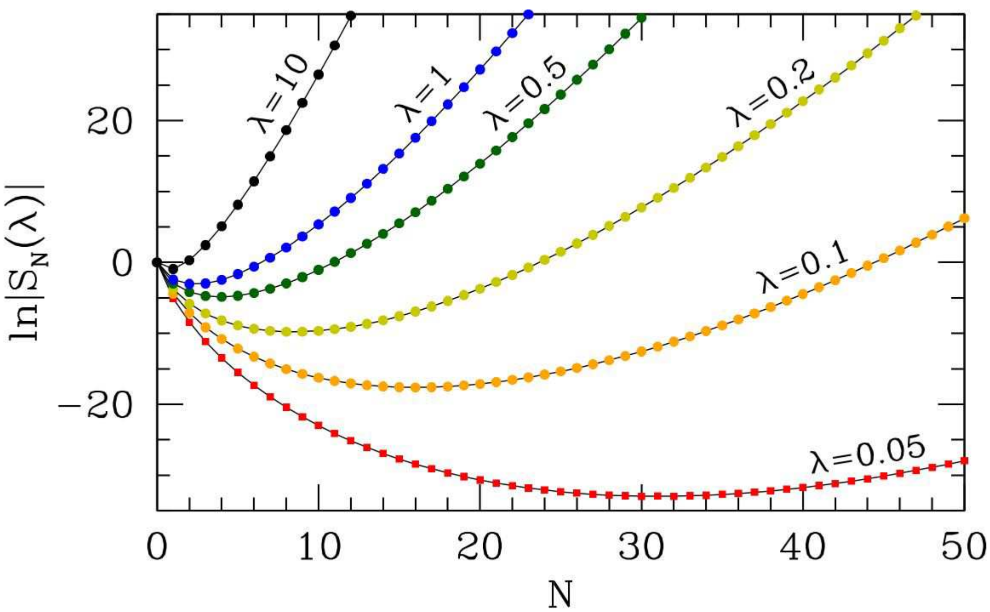


Figure 3: Relative error *versus* number of terms kept for the asymptotic series for $F(\lambda)$. Note that the optimal number of terms to sum is $N^*(\lambda) \approx \frac{3}{2\lambda}$.

suppose we have a diagram consisting of m_γ subdiagrams of type γ, for each γ. If the cluster γ contains n_γ vertices ($\times$), then we must have

$$N = \sum_\gamma m_\gamma \, n_\gamma \quad .$$

How many ways are there of assigning the labels to such a diagram? One might think $(4!)^N \cdot N!$, since for each vertex $\times$ there are $4!$ permutations of its four labels, and there are $N!$ ways to permute all the vertices. However, this overcounts diagrams which are *invariant* under one or more of these permutations. We define the *symmetry factor* s_γ of the (unlabeled) cluster γ as the number of permutations of the indices of a corresponding labeled cluster which result in the same contraction. We can also permute the m_γ identical disjoint clusters of type γ.

Examples of clusters and their corresponding symmetry factors are provided in fig. 4, for all diagrams with $n_\gamma \leq 3$. There is only one diagram with $n_\gamma = 1$, resembling . To obtain $s_\gamma = 8$, note that each of the circles can be separately rotated by an angle π about the long symmetry axis. In addition, the figure can undergo a planar rotation by π about an axis which runs through the sole vertex and is normal to the plane of the diagram. This results in $s_\gamma = 2 \cdot 2 \cdot 2 = 8$. For the cluster , there is one extra circle, so $s_\gamma = 2^4 = 16$. The third diagram in figure shows two vertices connected by four lines. Any of the $4!$ permutations of these lines results in the same diagram. In addition, we may reflect about the vertical symmetry axis, interchanging the vertices, to obtain another symmetry operation. Thus $s_\gamma = 2 \cdot 4! = 48$. One might ask why we don't also count the planar rotation by π as a symmetry operation. The answer is that it is equivalent to a combination of a reflection and a permutation, so it is not in fact a distinct symmetry operation. (If it were distinct, then s_γ would be 96.) Finally, consider the last diagram in the figure, which resembles a sausage with three links joined at the ends into a circle. If we keep the vertices fixed, there are 8 symmetry operations associated with the freedom to exchange the two lines associated with each of the three sausages. There are an additional 6 symmetry operations associated with permuting the three vertices, which can be classified as three in-plane rotations by 0, $\frac{2\pi}{3}$ and $\frac{4\pi}{3}$, each of which can also be combined with a reflection about the y-axis (this is known as the group C_{3v}). Thus, $s_\gamma = 8 \cdot 6 = 48$.

Now let us compute an expression for $F(\gamma)$ in terms of the clusters. We sum over all possible numbers of clusters

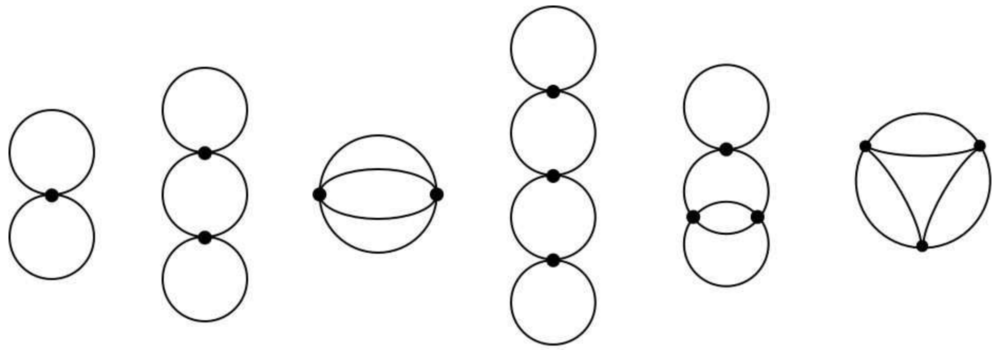


Figure 4: Cluster symmetry factors. A vertex is represented as a black dot (•) with four 'legs'.

at each order:

$$F(\gamma) = \sum_{N=0}^{\infty} \frac{1}{N!} \sum_{\{m_\gamma\}} \frac{(4!)^N N!}{\prod_\gamma s_\gamma^{m_\gamma}\, m_\gamma!} \left(-\frac{\lambda}{4!}\right)^N \delta_{N,\sum_\gamma m_\gamma n_\gamma}$$

$$= \exp\left(\sum_\gamma \frac{(-\lambda)^{n_\gamma}}{s_\gamma}\right) \quad .$$

Thus,

$$\ln F(\gamma) = \sum_\gamma \frac{(-\lambda)^{n_\gamma}}{s_\gamma} \quad ,$$

and *the logarithm of the sum over all diagrams is a sum over connected clusters*. It is instructive to work this out to order λ^2. We have, from the results of part (b),

$$F(\lambda) = 1 - \tfrac{1}{8}\,\lambda + \tfrac{35}{384}\,\lambda^2 + \mathcal{O}(\lambda^3) \quad \Longrightarrow \quad \ln F(\lambda) = -\tfrac{1}{8}\,\lambda + \tfrac{1}{12}\,\lambda^2 + \mathcal{O}(\lambda^3) \quad .$$

Note that there is one diagram with $N = 1$ vertex, with symmetry factor $s = 8$. For $N = 2$ vertices, there are two diagrams, one with $s = 16$ and one with $s = 48$ (see fig. 4). Since $\frac{1}{16} + \frac{1}{48} = \frac{1}{12}$, the diagrammatic expansion is verified to order λ^2.

(d) We now have[4]

$$G(\lambda) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\infty}\!dx\, \exp\left(-\frac{1}{2}\,x^2 - \frac{\lambda}{6!}\,x^6\right)$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!}\left(-\frac{\lambda}{6!}\right)^n \langle x^{6n}\rangle = \sum_{n=0}^{\infty} C_n\,\lambda^n \quad ,$$

where

$$C_n = \frac{(-1)^n\,(6n)!}{(6!)^n\, n!\, 2^{3n}\,(3n)!} \quad .$$

Invoking Stirling's approximation, we find

$$\ln |C_n| \sim 2n\,\ln n - \left(2 + \ln \tfrac{5}{3}\right) n \quad .$$

[4] According to `Mathematica`, the $G(\lambda)$ has the analytic form $G(\lambda) = \pi\sqrt{u}\left[\mathrm{Ai}^2(u) + \mathrm{Bi}^2(u)\right]$, where $u = (15/2\lambda)^{1/3}$ and $\mathrm{Ai}(z)$ and $\mathrm{Bi}(z)$ are Airy functions. The definitions and properties of the Airy functions are discussed in §9.2 of the NIST *Handbook of Mathematical Functions*.

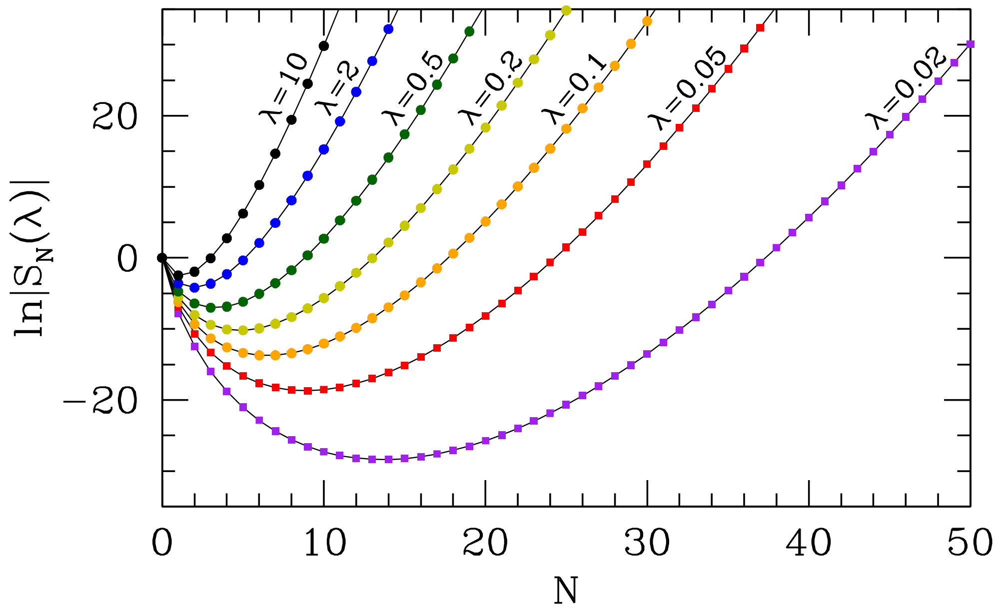


Figure 5: Logarithm of ratio of remainder after N terms $R_N(\lambda)$ to the value of the integral $G(\lambda)$, for various values of λ.

From the above expression for C_n, we see that the magnitude of the contribution of the $n^{\rm th}$ term in the perturbation series is

$$C_n\,\lambda^n = (-1)^n \exp\Big(2n\ln n - \big(2+\ln\tfrac{10}{3}\big)\,n + n\ln\lambda\Big) \quad .$$

Differentiating, we find that this contribution is minimized for $n=n^*(\lambda)$, where

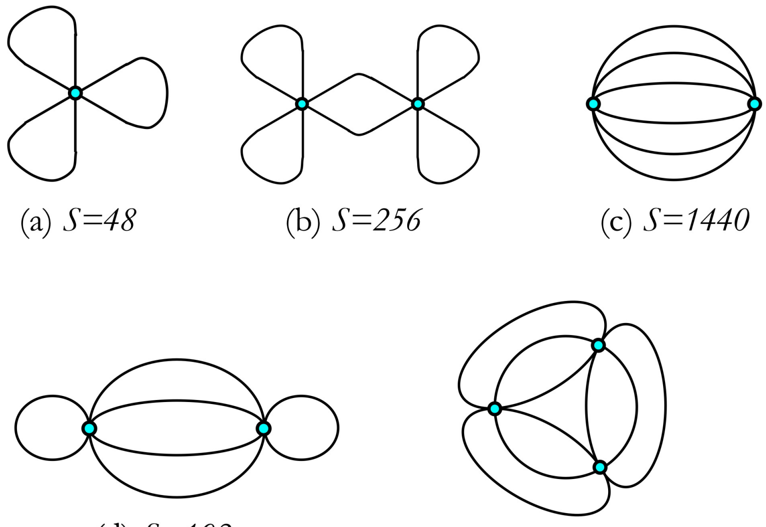
$$n^*(\lambda)=\sqrt{\frac{10}{3\lambda}} \quad .$$

Via numerical integration using FORTRAN subroutines from QUADPACK, one obtains the results in Fig. 5 and Tab. **??**.

λ	10	2	0.5	0.2	0.1	0.05	0.02
F	0.92344230	0.97298847	0.99119383	0.996153156	0.99800488	0.99898172	0.99958723
n^*	0.68	1.3	2.6	4.1	5.8	8.2	13

Table 2: $F(\lambda)$ and $n^*(\lambda)$ for problem 8d.

The series for $G(\lambda)$ and for $\ln G(\lambda)$ again have diagrammatic interpretations. If we expand the sextic term in a power series, each power of λ brings an additional six powers of x. It is natural to represent each such sextet with as a vertex with six legs. At order N of the series expansion, we have N such vertices and $6N$ legs to contract. As before, each full contraction of the leg indices may be represented as a labeled diagram, which is in general composed of several disjoint connected clusters. If the cluster γ contains n_γ vertices, then for any diagram we again must have $N=\sum_\gamma m_\gamma n_\gamma$, where m_γ is the number of times the cluster γ appears. As with the quartic example, the number of ways of assigning labels to a given diagram is given by the total number of possible permutations $(6!)^N\cdot N!$ divided by a correction factor $\prod_\gamma s_\gamma^{m_\gamma}\, m_\gamma!$, where s_γ is the symmetry factor of the cluster γ, and the $m_\gamma!$ term accounts for the possibility of permuting among different labeled clusters of the same type γ.



(a) S=48 (b) S=256 (c) S=1440

(d) S=192 (e) S=1296

Figure 6: Diagrams and their symmetry factors for the $\frac{1}{6!}\lambda\, x^6$ zero-dimensional field theory

Examples of clusters and their corresponding symmetry factors are provided in Fig. 6. There is only one diagram with $n_\gamma = 1$, shown panel (a), resembling a three-petaled flower. To obtain $s_\gamma = 48$, note that each of the petals can be rotated by 180° about an axis bisecting the petal, yielding a factor of 2^3. The three petals can then be permuted, yielding an additional factor of 3!. Hence the total symmetry factor is $s_\gamma = 2^3 \cdot 3! = 48$. Now we can see how dividing by the symmetry factor saves us from overcounting. In this case, we get $6!/s_\gamma = 720/48 = 15 = 5 \cdot 3 \cdot 1$, which is the correct number of contractions. For the diagram in panel (b), the four petals and the central loop can each be rotated about a symmetry axis, yielding a factor 2^5. The two left petals can be permuted, as can the two right petals. Finally, the two vertices can themselves be permuted. Thus, the symmetry factor is $s_\gamma = 2^5 \cdot 2^2 \cdot 2 = 2^8 = 256$. In panel (c), the six lines can be permuted (6!) and the vertices can be exchanged (2), hence $s_\gamma = 6! \cdot 2 = 1440$. In panel (d), the two outer loops each can be twisted by 180°, the central four lines can be permuted, and the vertices can be permuted, hence $s_\gamma = 2^2 \cdot 4! \cdot 2 = 192$. Finally, in panel (e), each pair of vertices is connected by three lines which can be permuted, and the vertices themselves can be permuted, so $s_\gamma = (3!)^3 \cdot 3! = 1296$.

Now let us compute an expression for $F(\gamma)$ in terms of the clusters. We sum over all possible numbers of clusters at each order:

$$
\begin{aligned}
G(\gamma) &= \sum_{N=0}^{\infty} \frac{1}{N!} \sum_{\{m_\gamma\}} \frac{(6!)^N N!}{\prod_\gamma s_\gamma^{m_\gamma}\, m_\gamma!} \left(-\frac{\lambda}{6!}\right)^N \delta_{N,\sum_\gamma m_\gamma n_\gamma} \\
&= \exp\left(\sum_\gamma \frac{(-\lambda)^{n_\gamma}}{s_\gamma}\right) \quad .
\end{aligned}
$$

Thus,

$$
\ln G(\gamma) = \sum_\gamma \frac{(-\lambda)^{n_\gamma}}{s_\gamma} \quad ,
$$

and *the logarithm of the sum over all diagrams is a sum over connected clusters*. It is instructive to work this out to order λ^2. We have, from the results of part (a),

$$
G(\lambda) = 1 - \frac{\lambda}{2^6 \cdot 3} + \frac{7\cdot 11\cdot \lambda^2}{2^9\cdot 3\cdot 5} + \mathcal{O}(\lambda^3) \quad \Longrightarrow \quad \ln G(\lambda) = -\frac{\lambda}{2^6\cdot 3} + \frac{113\cdot\lambda^2}{2^8\cdot 3^2\cdot 5} + \mathcal{O}(\lambda^3) \quad .
$$

Note that there is one diagram with $N = 1$ vertex, with symmetry factor $s = 48$. For $N = 2$ vertices, there are three diagrams, one with $s = 256$, one with $s = 1440$, and one with $s = 192$ (see Fig. 6). Since $\frac{1}{256} + \frac{1}{1440} + \frac{1}{192} = \frac{113}{2^8 3^2 5}$, the diagrammatic expansion is verified to order λ^2.

(e) In quantum field theory (QFT), the vertices themselves carry space-time labels, and the contractions, *i.e.* the lines connecting the legs of the vertices, are *propagators* $G(x_i^\mu - x_j^\mu)$, where x_i^μ is the space-time label associated with vertex i. It is convenient to work in momentum-frequency space, in which case we work with the Fourier transform $\hat{G}(p^\mu)$ of the space-time propagators. Integrating over the space-time coordinates of each vertex then enforces total 4-momentum conservation at each vertex. We then must integrate over all the internal 4-momenta to obtain the numerical value for a given diagram. The diagrams, as you know, are associated with Feynman's approach to QFT and are known as Feynman diagrams. Our example here is equivalent to a $(0 + 0)$-dimensional field theory, *i.e.* zero space dimensions and zero time dimensions. There are then no internal 4-momenta to integrate over, and each propagator is simply a number rather than a function. The discussion above of symmetry factors s_γ carries over to the more general QFT case.

There is an important lesson to be learned here about the behavior of asymptotic series. As we have seen, if λ is sufficiently small, summing more and more terms in the perturbation series results in better and better results, until one reaches an optimal order when the error is minimized. Beyond this point, summing additional terms makes the result *worse*, and indeed the perturbation series diverges badly as $N \to \infty$. Typically the optimal order of perturbation theory is inversely proportional to the coupling constant. For quantum electrodynamics (QED), where the coupling constant is the fine structure constant $\alpha = e^2/\hbar c \approx \frac{1}{137}$, we lose the ability to calculate in a reasonable time long before we get to 137 loops, so practically speaking no problems arise from the lack of convergence. In quantum chromodynamics (QCD), however, the effective coupling constant is about two orders of magnitude larger, and perturbation theory is a much more subtle affair.