2

Calculation of Expectations

2.1 Introduction

Many of the hardest problems in applied probability revolve around the calculation of expectations of one sort or another. On one level, these are merely humble exercises in integration or summation. However, we should not be so quick to dismiss the intellectual challenges. Readers are doubtless already aware of the clever applications of characteristic and moment generating functions. This chapter is intended to review and extend some of the tools that probabilists routinely call on. Readers can consult the books [34, 59, 60, 78, 80, 166] for many additional examples of these tools in action.

2.2 Indicator Random Variables and Symmetry

Many counting random variables can be expressed as the sum of indicator random variables. If $S = \sum_{i=1}^{n} 1_{A_i}$ for events $A_1, \ldots, A_n$, then straightforward calculations and equation (1.10) give

$$\mathrm{E}(S) = \sum_{i=1}^{n} \Pr(A_i) \tag{2.1}$$

$$\mathrm{Var}(S) = \sum_{i=1}^{n} \Pr(A_i) + \sum_{i=1}^{n} \sum_{j \neq i} \Pr(A_i \cap A_j) - \mathrm{E}(S)^2. \tag{2.2}$$

K. Lange, *Applied Probability*,
DOI 10.1007/978-1-4419-7165-4_2,

Example 2.2.1 *Fixed Points of a Random Permutation*

There are $n!$ permutations π of the set $\{1, \ldots, n\}$. Under the uniform distribution, each of these permutations is equally likely. If A_i is the event that $\pi(i) = i$, then $S = \sum_{i=1}^n 1_{A_i}$ is the number of fixed points of π. By symmetry, $\Pr(A_i) = \frac{1}{n}$ and

$$\begin{aligned}\Pr(A_i \cap A_j) &= \Pr(A_j \mid A_i)\Pr(A_i) \\ &= \frac{1}{(n-1)n}.\end{aligned}$$

Hence, the formulas in (2.2) yield $\mathrm{E}(S) = \frac{n}{n} = 1$ and

$$\begin{aligned}\mathrm{Var}(S) &= \frac{n}{n} + \sum_{i=1}^n \sum_{j \neq i} \frac{1}{(n-1)n} - 1^2 \\ &= 1.\end{aligned}$$

The equality $\mathrm{E}(S) = \mathrm{Var}(S)$ suggests that S is approximately Poisson distributed. We will verify this conjecture in Example 4.3.1. ■

Example 2.2.2 *Pattern Matching*

Consider a random string of n letters drawn uniformly and independently from the alphabet $\{1, \ldots, m\}$. Let S equal the number of occurrences of a given word of length $k \leq n$ in the string. For example, with $m = 2$ and $n = 10$, all strings have probability 2^{-10}. The word 101 is present in the string 1101011101 three times. Represent S as

$$S = \sum_{j=1}^{n-k+1} 1_{A_j},$$

where A_j is the event that the given word occurs beginning at position j in the string. In view of equation (2.1), it is obvious that

$$\mathrm{E}(S) = \sum_{j=1}^{n-k+1} \Pr(A_j) = (n-k+1)p^k$$

for the choice $p = m^{-1}$. Calculation of $\mathrm{Var}(S)$ is more subtle. Equation (2.2) and symmetry imply

$$\mathrm{Var}(S) = (n-k+1)\,\mathrm{Var}(1_{A_1}) + 2\sum_{j=2}^{l}(n-k-j+2)\,\mathrm{Cov}(1_{A_1}, 1_{A_j}),$$

where $l = \min\{k, n-k+1\}$, the multiplier $2(n-k-j+2)$ of $\mathrm{Cov}(1_{A_1}, 1_{A_j})$ equals the number of pairs (r, x) with $|r-s| = j-1$, and the events A_r and A_s are independent whenever $|r-s| \geq k$. Although it is clear that

$$\mathrm{Var}(1_{A_1}) = p^k - p^{2k},$$

the covariance terms present more of a challenge because of the possibility of overlapping occurrences of the word. Let ϵ_l equal 1 or 0, depending on whether the last l letters of the word taken as a block coincide with the first l letters of the word taken as a block. For the particular word 101, $\epsilon = 1$ and $\epsilon_2 = 0$. With this convention, we calculate

$$\mathrm{Cov}(1_{A_1}, 1_{A_j}) \quad = \quad \epsilon_{k-j+1} p^{k+j-1} - p^{2k}$$

for $2 \leq j \leq k$. ■

Both of the previous examples exploit symmetry as well as indicator random variables. Here is another example from sampling theory that depends crucially on symmetry [40].

Example 2.2.3 *Sampling without Replacement*

Assume that m numbers $Y_1, \ldots, Y_m$ are drawn randomly without replacement from n numbers $x_1, \ldots, x_n$. It is of interest to calculate the mean and variance of the sample average $S = \frac{1}{m}\sum_{i=1}^m Y_i$. Clearly,

$$\mathrm{E}(S) \quad = \quad \frac{1}{m}\sum_{i=1}^m \mathrm{E}(Y_i) \quad = \quad \bar{x},$$

where $\bar{x}$ is the sample average of the x_i. To calculate the variance of S, let $s^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$ denote the sample variance of the x_i. Now imagine filling out the sample to $Y_1, \ldots, Y_n$ so that all n values $x_1, \ldots, x_n$ are exhausted. Because the sum $Y_1 + \cdots + Y_n = n\bar{x}$ is constant, symmetry and equation (1.10) imply that

$$\begin{aligned} 0 &= \mathrm{Var}(Y_1 + \cdots + Y_n) \\ &= ns^2 + n(n-1)\,\mathrm{Cov}(Y_1, Y_2). \end{aligned}$$

In verifying that $\mathrm{Cov}(Y_i, Y_j) = \mathrm{Cov}(Y_1, Y_2)$, it is helpful to think of the sampling being done simultaneously rather than sequentially. In any case, $\mathrm{Cov}(Y_1, Y_2) = -\frac{s^2}{n-1}$, and the formula

$$\begin{aligned} \mathrm{Var}(S) &= \frac{1}{m^2}\Big[ms^2 + m(m-1)\,\mathrm{Cov}(Y_1, Y_2)\Big] \\ &= \frac{1}{m^2}\left[ms^2 - \frac{m(m-1)s^2}{n-1}\right] \\ &= \frac{(n-m)s^2}{m(n-1)} \end{aligned}$$

follows directly. ■

The next problem, the first of a long line of problems in geometric probability, also yields to symmetry arguments [116].

Example 2.2.4 *Buffon Needle Problem*

Suppose we draw an infinite number of equally distant parallel lines on the plane R^2. If we drop a needle (or line segment) of fixed length randomly onto the plane, then the needle may or may not intersect one of the parallel lines. Figure 2.1 shows the needle intersecting a line. Buffon's problem is to calculate the probability of an intersection. Without loss of generality, we assume that the spacing between lines is 1 and the length of the needle is d. Let X_d be the random number of lines that the needle intersects. If $d < 1$, then X_d equals 0 or 1, and $\Pr(X_d = 1) = \mathrm{E}(X_d)$. Thus, Buffon's problem reduces to calculating an expectation for a short needle. Our task is to construct the function $f(d) = \mathrm{E}(X_d)$. This function is clearly nonnegative, increasing, and continuous in d.

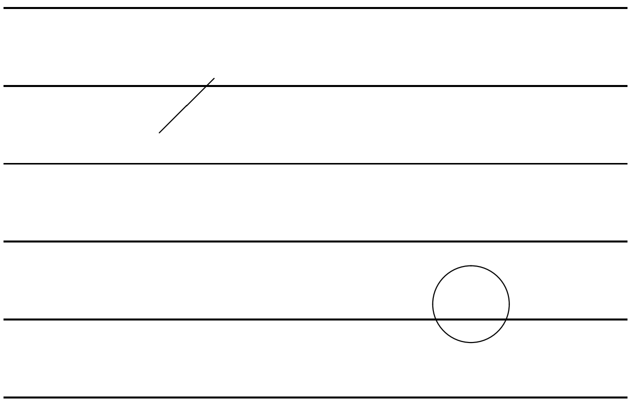

FIGURE 2.1. Diagram of the Buffon Needle Problem

Now imagine randomly dropping two needles simultaneously of lengths d_1 and d_2. The expected number of intersections of both needles obviously amounts to $\mathrm{E}(X_{d_1}) + \mathrm{E}(X_{d_2})$. This result holds whether we drop the two needles independently or dependently, as long as we drop them randomly. We can achieve total dependence by welding the end of one needle to the start of the other needle. If the weld is just right, then the two needles will form a single needle of length $d_1 + d_2$. This shows that

$$f(d_1 + d_2) = f(d_1) + f(d_2). \tag{2.3}$$

The only functions $f(d)$ that are nonnegative, increasing, and additive in d are the linear functions $f(d) = cd$ with $c \geq 0$. To find the proportionality constant c, we take the experiment of welding needles together to its logical extreme. Thus, a rigid wire of welded needles with perimeter p determines on average cp intersections. In the limit, we can replace the wire by any reasonable curve. The key to finding c is to take a circle of diameter 1. This particular curve has perimeter π and either is tangent to two lines or intersects the same line twice. Figure 2.1 depicts the latter case. The equation $2 = c\pi$ now determines $c = 2/\pi$ and $f(d) = 2d/\pi$. ■

2.3 Conditioning

A third way to calculate expectations is to condition. Two of the next three examples use conditioning to derive a recurrence relation. In the family planning model, the recurrence is difficult to solve exactly, but as with most recurrences, it is easy to implement by hand or computer.

Example 2.3.1 *Beta-Binomial Distribution*

Consider a random variable P with beta density

$$f_{\alpha\beta}(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

on the unit interval. In Section 2.9, we generalize the beta distribution to the Dirichlet distribution. In the meantime, the reader may recall the moment calculation

$$\begin{aligned}
\mathrm{E}[P^i(1-P)^j] &= \int_0^1 p^i(1-p)^j f_{\alpha\beta}(p)\,dp \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+i)\Gamma(\beta+j)}{\Gamma(\alpha+\beta+i+j)} \int_0^1 f_{\alpha+i,\beta+j}(p)\,dp \\
&= \frac{(\alpha+i-1)\cdots\alpha(\beta+j-1)\cdots\beta}{(\alpha+\beta+i+j-1)\cdots(\alpha+\beta)}
\end{aligned}$$

invoking the factorial property $\Gamma(x+1) = x\Gamma(x)$ of the gamma function. This gives, for example,

$$\begin{aligned}
\mathrm{E}(P) &= \frac{\alpha}{\alpha+\beta} \\
\mathrm{E}[P(1-P)] &= \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)} \qquad (2.4)\\
\mathrm{Var}(P) &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.
\end{aligned}$$

Now suppose we carry out n Bernoulli trials with the same random success probability P pertaining to all n trials. The number of successes S_n follows a beta-binomial distribution. Application of equations (1.3) and

(1.11) yields

$$\begin{aligned}
\mathrm{E}(S_n) &= n\,\mathrm{E}(P) \\
\mathrm{Var}(S_n) &= \mathrm{E}[\mathrm{Var}(S_n \mid P)] + \mathrm{Var}[\mathrm{E}(S_n \mid P)] \\
&= \mathrm{E}[nP(1-P)] + \mathrm{Var}(nP) \\
&= n\,\mathrm{E}[P(1-P)] + n^2\,\mathrm{Var}(P),
\end{aligned}$$

which can be explicitly evaluated using the moments in equation (2.4). Problem 4 provides the density of S_n. ■

Example 2.3.2 *Repeated Uniform Sampling*

Suppose we construct a sequence of dependent random variables X_n by taking $X_0 = 1$ and sampling X_n uniformly from the interval $[0, X_{n-1}]$. To calculate the moments of X_n, we use the facts $\mathrm{E}(X_n^k) = \mathrm{E}[\mathrm{E}(X_n^k \mid X_{n-1})]$ and

$$\begin{aligned}
\mathrm{E}(X_n^k \mid X_{n-1}) &= \frac{1}{X_{n-1}} \int_0^{X_{n-1}} x^k\,dx \\
&= \left. \frac{x^{k+1}}{X_{n-1}(k+1)} \right|_0^{X_{n-1}} \\
&= \frac{1}{k+1} X_{n-1}^k.
\end{aligned}$$

Hence,

$$\mathrm{E}(X_n^k) = \frac{1}{k+1}\,\mathrm{E}(X_{n-1}^k) = \left(\frac{1}{k+1}\right)^n.$$

For example, $\mathrm{E}(X_n) = 2^{-n}$ and $\mathrm{Var}(X_n) = 3^{-n} - 2^{-2n}$. It is interesting that if we standardize by defining $Y_n = 2^n X_n$, then the mean $\mathrm{E}(Y_n) = 1$ is stable, but the variance $\mathrm{Var}(Y_n) = (\frac{4}{3})^n - 1$ tends to ∞.

The clouds of mystery lift a little when we rewrite X_n as the product

$$X_n = U_n X_{n-1} = \prod_{i=1}^n U_i$$

of independent uniform random variables $U_1, \ldots, U_n$ on $[0, 1]$. The product rule for expectations now gives $\mathrm{E}(X_n^k) = \mathrm{E}(U_1^k)^n = (k+1)^{-n}$. Although we cannot stabilize X_n, it is possible to stabilize $\ln X_n$. Indeed, Problem 5 notes that $\ln X_n = \sum_{i=1}^n \ln U_i$ follows a $-\frac{1}{2}\chi^2_{2n}$ distribution with mean $-n$ and variance n. Thus for large n, the central limit theorem implies that $(\ln X_n + n)/\sqrt{n}$ has an approximate standard normal distribution. ■

Example 2.3.3 *Expected Family Size*

A married couple desires a family consisting of at least s sons and d daughters. At each birth, the mother independently bears a son with probability p and a daughter with probability $q = 1-p$. They will quit having children when their objective is reached. Let N_{sd} be the random number of children born to them. Suppose we wish to calculate the expected value $\mathrm{E}(N_{sd})$. Two cases are trivial. If either $s = 0$ or $d = 0$, then N_{sd} follows a negative binomial distribution. Therefore, $\mathrm{E}(N_{0d}) = d/q$ and $\mathrm{E}(N_{s0}) = s/p$. When both s and d are positive, the distribution of N_{sd} is not so obvious. Conditional on the sex of the first child, the random variable $N_{sd} - 1$ is either a probabilistic copy $N^*_{s-1,d}$ of $N_{s-1,d}$ or a probabilistic copy $N^*_{s,d-1}$ of $N_{s,d-1}$. Because in both cases the copy is independent of the sex of the first child, the recurrence relation

$$\begin{aligned}\mathrm{E}(N_{sd}) &= p[1+\mathrm{E}(N_{s-1,d})] + q[1+\mathrm{E}(N_{s,d-1})] \\ &= 1 + p\,\mathrm{E}(N_{s-1,d}) + q\,\mathrm{E}(N_{s,d-1})\end{aligned}$$

follows from conditioning on this outcome.

There are many variations on this idea. For instance, suppose we wish to compute the probability R_{sd} that they reach their quota of s sons before their quota of d daughters. Then the R_{sd} satisfy the boundary conditions $R_{0d} = 1$ for $d > 0$ and $R_{s0} = 0$ for $s > 0$. When s and d are both positive, we have the recurrence relation

$$R_{sd} = pR_{s-1,d} + qR_{s,d-1}.$$

2.4 Moment Transforms

Each of the moment transforms reviewed in Section 1.5 can be differentiated to capture the moments of a random variable. Equally important, these transforms often solve other theoretical problems with surprising ease. The next seven examples illustrate these two roles.

Example 2.4.1 *Characteristic Function of a Normal Density*

To find the characteristic function $\hat{\psi}(t) = \mathrm{E}(e^{itX})$ of a standard normal random variable X with density $\psi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, we derive and solve a differential equation. Differentiation under the integral sign and integration by parts together imply that

$$\frac{d}{dt}\hat{\psi}(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itx} ixe^{-\frac{x^2}{2}}\,dx$$

$$\begin{aligned}
&= -\frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itx}\frac{d}{dx}e^{-\frac{x^2}{2}}\,dx \\
&= \frac{-i}{\sqrt{2\pi}}e^{itx}e^{-\frac{x^2}{2}}\Big|_{-\infty}^{\infty} - \frac{t}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itx}e^{-\frac{x^2}{2}}\,dx \\
&= -t\hat{\psi}(t).
\end{aligned}$$

The unique solution to this differential equation with initial value $\hat{\psi}(0) = 1$ is $\hat{\psi}(t) = e^{-t^2/2}$.

If X is a standard normal random variable, then $\mu + \sigma X$ is a normal random variable with mean μ and variance σ^2. The general identity $\mathrm{E}[e^{it(\mu+\sigma X)}] = e^{it\mu}\,\mathrm{E}[e^{i(\sigma t)X}]$ permits us to express the characteristic function of the normal distribution with mean μ and variance σ^2 as

$$\hat{\psi}_{\mu,\sigma^2}(t) \quad = e^{it\mu}\hat{\psi}(\sigma t) \quad = \quad e^{it\mu - \frac{\sigma^2 t^2}{2}}.$$

The first two derivatives

$$\begin{aligned}
\frac{d}{dt}\hat{\psi}_{\mu,\sigma^2}(t) &= (i\mu - \sigma^2 t)e^{it\mu - \frac{\sigma^2 t^2}{2}} \\
\frac{d^2}{dt^2}\hat{\psi}_{\mu,\sigma^2}(t) &= -\sigma^2 e^{it\mu - \frac{\sigma^2 t^2}{2}} + (i\mu - \sigma^2 t)^2 e^{it\mu - \frac{\sigma^2 t^2}{2}}
\end{aligned}$$

evaluated at $t = 0$ determine the mean μ and second moment $\sigma^2 + \mu^2$ as indicated in equation (1.8). ■

Example 2.4.2 *Characteristic Function of a Gamma Density*

A random variable X with exponential density $\lambda e^{-\lambda x}1_{\{x>0\}}$ has characteristic function

$$\begin{aligned}
\int_0^{\infty} e^{itx}\lambda e^{-\lambda x}\,dx &= \frac{\lambda}{it - \lambda}e^{(it-\lambda)x}\Big|_0^{\infty} \\
&= \frac{\lambda}{\lambda - it}.
\end{aligned}$$

An analogous calculation yields the Laplace transform $\lambda/(\lambda + t)$. Differentiation of either of these transforms produces

$$\begin{aligned}
\mathrm{E}(X) &= \frac{1}{\lambda} \\
\mathrm{Var}(X) &= \frac{1}{\lambda^2}.
\end{aligned}$$

The gamma density $\lambda^n x^{n-1}e^{-\lambda x}1_{\{x>0\}}/\Gamma(n)$ is the convolution of n exponential densities with common intensity λ. The corresponding random variable X_n therefore has

$$\mathrm{E}(X_n) \quad = \quad \frac{n}{\lambda}$$

$$\begin{aligned}
\mathrm{Var}(X_n) &= \frac{n}{\lambda^2} \\
\mathrm{E}(e^{itX_n}) &= \Big(\frac{\lambda}{\lambda - it}\Big)^n \\
\mathrm{E}(e^{-tX_n}) &= \Big(\frac{\lambda}{\lambda + t}\Big)^n.
\end{aligned}$$

Problem 16 indicates that these results carry over to non-integer $n > 0$. ■

Example 2.4.3 *Factorial Moments*

Let X be a nonnegative, integer-valued random variable. Repeated differentiation of its probability generating function $G(u) = \mathrm{E}(u^X)$ yields its factorial moments $\mathrm{E}[X(X-1)\cdots(X-j+1)] = \frac{d^j}{du^j}G(1)$. The first two central moments

$$\begin{aligned}
\mathrm{E}(X) &= G'(1) \\
\mathrm{Var}(X) &= \mathrm{E}[X(X-1)] + \mathrm{E}(X) - \mathrm{E}(X)^2 \\
&= G''(1) + G'(1) - G'(1)^2
\end{aligned}$$

are worth committing to memory. As an example, suppose X is Poisson distributed with mean λ. Then

$$G(u) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} u^k = e^{-\lambda(1-u)}.$$

Repeated differentiation yields $\frac{d^j}{du^j}G(1) = \lambda^j$. In particular, $\mathrm{E}(X) = \lambda$ and $\mathrm{Var}(X) = \lambda$. For another example, let X follow a binomial distribution with n trials and success probability p per trial. In this case $G(u) = (1-p+pu)^n$, $\mathrm{E}(X) = np$, and $\mathrm{Var}(X) = np(1-p)$. ■

Example 2.4.4 *Random Sums*

Suppose $X_1, X_2, \ldots$ is a sequence of independent identically distributed (i.i.d.) random variables. Consider the random sum $S_N = \sum_{i=1}^{N} X_i$, where the random number of terms N is independent of the X_i, and where we adopt the convention $S_0 = 0$. For example in an ecological study, the number of animal litters N in a plot of land might have a Poisson distribution to a good approximation. The random variable X_i then represents the number of offspring in litter i, and the compound Poisson random variable S_N counts the number of offspring over the whole plot.

If N has probability generating function $G(u) = \mathrm{E}(u^N)$, then the characteristic function of S_N is

$$\mathrm{E}(e^{itS_N}) = \sum_{n=0}^{\infty} \mathrm{E}(e^{itS_N} \mid N = n) \Pr(N = n)$$

$$
\begin{aligned}
&= \sum_{n=0}^{\infty} \mathrm{E}(e^{itX_1})^n \Pr(N = n) \\
&= G[\mathrm{E}(e^{itX_1})].
\end{aligned}
$$

This composition rule carries over to other moment transforms. For instance, if the X_i are nonnegative and integer-valued with probability generating function $H(u)$, then a similar argument gives $\mathrm{E}(u^{S_N}) = G[H(u)]$.

We can extract the moments of S_N by differentiation. Alternatively, conditioning on N produces

$$
\mathrm{E}(S_N) = \mathrm{E}[N\,\mathrm{E}(X_1)] = \mathrm{E}(N)\,\mathrm{E}(X_1)
$$

and

$$
\begin{aligned}
\mathrm{Var}(S_N) &= \mathrm{E}[\mathrm{Var}(S_N \mid N)] + \mathrm{Var}[\mathrm{E}(S_N \mid N)] \\
&= \mathrm{E}[N\,\mathrm{Var}(X_1)] + \mathrm{Var}[N\,\mathrm{E}(X_1)] \\
&= \mathrm{E}(N)\,\mathrm{Var}(X_1) + \mathrm{Var}(N)\,\mathrm{E}(X_1)^2.
\end{aligned}
$$

For instance, if N has a Poisson distribution with mean λ and the X_i have a binomial distribution with parameters n and p, then S_N has

$$
\begin{aligned}
\mathrm{E}(u^{S_N}) &= e^{-\lambda[1-(1-p+pu)^n]} \\
\mathrm{E}(S_N) &= \lambda np \\
\mathrm{Var}(S_N) &= \lambda np(1-p) + \lambda n^2 p^2
\end{aligned}
$$

as its probability generating function, mean, and variance, respectively. ■

Example 2.4.5 *Sum of Uniforms*

In Example 2.3.2, we considered the product of n independent random variables $U_1, \ldots, U_n$ uniformly distributed on $[0, 1]$. We now turn to the problem of finding the density of the sum $S_n = U_1 + \cdots + U_n$. Our strategy will be to calculate and invert the Laplace transform of the density of S_n, keeping in mind that the Laplace transform of a random variable coincides with the Laplace transform of its density. Because the Laplace transform of a single U_i is $\int_0^1 e^{-tx}dx = (1 - e^{-t})/t$, the Laplace transform of S_n is

$$
\frac{(1-e^{-t})^n}{t^n} = \frac{1}{t^n}\sum_{k=0}^{n}\binom{n}{k}(-1)^k e^{-kt}.
$$

In view of the linearity of the Laplace transform, it therefore suffices to invert the term e^{-kt}/t^n. Since multiplication by e^{-kt} in the transform domain corresponds to an argument shift of k in the original domain, all we need to do is find the function with transform t^{-n} and shift it by k. We now make an inspired guess that the function x^{n-1} is relevant. Because the Laplace transform deals with functions defined on $[0, \infty)$, we exchange x^{n-1} for the function $(x)_+^{n-1}$, which equals 0 for $x \leq 0$ and x^{n-1} for $x > 0$. The change of variables $u = tx$ and the definition of the gamma function show that $(x)_+^{n-1}$ has transform

$$\begin{aligned} \int_0^\infty x^{n-1} e^{-tx} dx &= \frac{1}{t^n} \int_0^\infty u^{n-1} e^{-u} du \\ &= \frac{(n-1)!}{t^n}. \end{aligned}$$

Up to a constant, this is just what we need. Hence, we conclude that S_n has density

$$f(x) = \frac{1}{(n-1)!} \sum_{k=0}^{n} \binom{n}{k} (-1)^k (x-k)_+^{n-1}.$$

The corresponding distribution function

$$F(x) = \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} (-1)^k (x-k)_+^{n}$$

emerges after integration with respect to x. ■

Example 2.4.6 *A Nonexistence Problem*

Is it always possible to represent a random variable X as the difference $Y - Z$ of two independent, identically distributed random variables Y and Z? The answer is clearly no unless X is symmetrically distributed around 0. For a symmetrically distributed X, the question is more subtle. Suppose that Y and Z exist for such an X. Then the characteristic function of X reduces to

$$\begin{aligned} \mathrm{E}[e^{it(Y-Z)}] &= \mathrm{E}(e^{itY})\,\mathrm{E}(e^{-itZ}) \\ &= \mathrm{E}(e^{itY})\,\mathrm{E}(e^{itY})^* \\ &= |\,\mathrm{E}(e^{itY})|^2, \end{aligned}$$

where the superscript $*$ denotes complex conjugation. (It is trivial to check that conjugation commutes with expectation for complex random variables possessing only a finite number of values, and this property persists in the limit for all complex random variables.) In any case, if the representation $X = Y - Z$ holds, then the characteristic function of X is nonnegative. Thus, to construct a counterexample, all we need to do is find a symmetrically distributed random variable whose characteristic function fails the test of nonnegativity. For instance, if we take X to be uniformly distributed on $[-\frac{1}{2}, \frac{1}{2}]$, then its characteristic function

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{itx} dx = \left.\frac{e^{itx}}{it}\right|_{-\frac{1}{2}}^{\frac{1}{2}} = \frac{\sin(\frac{t}{2})}{\frac{t}{2}}$$

oscillates in sign. ■

Example 2.4.7 *Characterization of the Standard Normal Distribution*

Consider a random variable X with mean 0, variance 1, and characteristic function $\hat{\psi}(t)$. If X is standard normal and Y is an independent copy of X, then for all a and b the sum $aX + bY$ has the same distribution as $\sqrt{a^2+b^2}X$. This distributional identity implies the characteristic function identity

$$\hat{\psi}(at)\hat{\psi}(bt) \quad = \quad \hat{\psi}\left(\sqrt{a^2+b^2}t\right) \tag{2.5}$$

for all t.

Conversely, suppose the functional equation (2.5) holds for a random variable X with mean 0 and variance 1. Let us show that X possesses a standard normal distribution. The special case $a = -1$ and $b = 0$ of equation (2.5) amounts to $\hat{\psi}(-t) = \hat{\psi}(t)$, from which it immediately follows that $\hat{\psi}(t)^* = \hat{\psi}(-t) = \hat{\psi}(t)$. Thus, $\hat{\psi}(t)$ is real and even. It is also differentiable because $\mathrm{E}(|X|) < \infty$. Now define $g(t^2) = \hat{\psi}(t)$ for $t > 0$. Setting $t = 1$ and replacing a^2 by a and b^2 by b in the functional equation (2.5) produce the revised functional equation

$$g(a)g(b) \quad = \quad g(a+b). \tag{2.6}$$

Taking $f(t) = \ln g(t)$ lands us right back at equation (2.3), except that it is no longer clear that $f(t)$ is monotone. Rather than rely on our previous hand-waving solution, we can differentiate equation (2.6), first with respect to $a \geq 0$ and then with respect to $b \geq 0$. This yields

$$g'(a)g(b) \quad = \quad g'(a+b) \quad = \quad g(a)g'(b). \tag{2.7}$$

If we take $b > 0$ sufficiently small, then $g(b) > 0$, owing to the continuity of $g(t)$ and the initial condition $g(0) = 1$. Dividing equation (2.7) by $g(b)$ and defining $\lambda = -g'(b)/g(b)$ leads to the differential equation $g'(a) = -\lambda g(a)$ with solution $g(a) = e^{-\lambda a}$. To determine λ, note that the equality

$$g''(t^2)4t^2 + g'(t^2)2 \quad = \quad \hat{\psi}''(t)$$

yields $-2\lambda = -1$ in the limit as t approaches 0. Thus, $\hat{\psi}(t) = e^{-t^2/2}$ as required. ■

2.5 Tail Probability Methods

Consider a nonnegative random variable X with distribution function $F(x)$. The right-tail probability $\Pr(X > t) = 1 - F(t)$ turns out to be helpful in calculating certain expectations relative to X. Let $h(t)$ be an integrable function on each finite interval $[0, x]$. If we define $H(x) = H(0) + \int_0^x h(t)\,dt$ and suppose that $\int_0^\infty |h(t)|[1-F(t)]\,dt < \infty$, then Fubini's theorem justifies the calculation

$$
\begin{aligned}
\mathrm{E}[H(X)] &= H(0) + \mathrm{E}\left[\int_0^X h(t)\,dt\right] \\
&= H(0) + \int_0^\infty \int_0^x h(t)\,dt\,dF(x) \\
&= H(0) + \int_0^\infty \int_t^\infty dF(x)h(t)\,dt \qquad (2.8) \\
&= H(0) + \int_0^\infty h(t)[1 - F(t)]\,dt.
\end{aligned}
$$

If X is concentrated on the integers $\{0, 1, 2, \ldots\}$, the right-tail probability $1 - F(t)$ is constant except for jumps at these integers. Equation (2.8) therefore reduces to

$$
\mathrm{E}[H(X)] = H(0) + \sum_{k=0}^{\infty} [H(k+1) - H(k)][1 - F(k)]. \qquad (2.9)
$$

Example 2.5.1 *Moments from Right-Tail Probabilities*

The choices $h(t) = nt^{n-1}$ and $H(0) = 0$ yield $H(x) = x^n$. Hence, equations (2.8) and (2.9) become

$$
\mathrm{E}[X^n] = n\int_0^\infty t^{n-1}[1 - F(t)]\,dt
$$

and

$$
\mathrm{E}[X^n] = \sum_{k=0}^{\infty} [(k+1)^n - k^n][1 - F(k)],
$$

respectively. For instance, if X is exponentially distributed, then the right-tail probability $1 - F(t) = e^{-\lambda t}$ and $E(X) = \int_0^\infty e^{-\lambda t} dt = \lambda^{-1}$. If X is geometrically distributed with failure probability q, then $1 - F(k) = q^k$ and $\mathrm{E}(X) = \sum_{k=0}^\infty q^k = (1-q)^{-1}$. ■

Example 2.5.2 *Laplace Transforms*

Equation (2.8) also determines the relationship between the Laplace transform $\mathrm{E}(e^{-sX})$ of a nonnegative random variable X and the ordinary Laplace transform $\tilde{F}(s)$ of its distribution function $F(x)$. For this purpose, we choose $h(t) = -se^{-st}$ and $H(0) = 1$. The resulting integral $H(x) = e^{-sx}$

and equation (2.8) together yield the formula

$$\begin{aligned} \mathrm{E}(e^{-sX}) &= 1 - s\int_0^\infty e^{-st}[1-F(t)]\,dt \\ &= s\int_0^\infty e^{-st}F(t)\,dt \\ &= s\tilde{F}(s). \end{aligned}$$

For example, if X is exponentially distributed with intensity λ, then the Laplace transform $\mathrm{E}(e^{-sX}) = \lambda/(s+\lambda)$ mentioned in Example 2.4.2 leads to $\tilde{F}(s) = \lambda/[s(s+\lambda)]$. ■

2.6 Moments of Reciprocals and Ratios

Ordinarily we differentiate Laplace transforms to recover moments. However, to recover an inverse moment, we need to integrate [43]. Suppose X is a positive random variable with Laplace transform $L(t)$. If $n > 0$, then Fubini's theorem and the change of variables $s = tX$ shows that

$$\begin{aligned} \int_0^\infty t^{n-1}L(t)\,dt &= \mathrm{E}\left(\int_0^\infty t^{n-1}e^{-tX}dt\right) \\ &= \mathrm{E}\left(X^{-n}\int_0^\infty s^{n-1}e^{-s}ds\right) \\ &= \mathrm{E}\left(X^{-n}\right)\Gamma(n). \end{aligned}$$

The formula

$$\mathrm{E}\left(X^{-n}\right) = \frac{1}{\Gamma(n)}\int_0^\infty t^{n-1}L(t)\,dt \tag{2.10}$$

can be evaluated exactly in some cases. In other cases, for instance when n fails to be an integer, the formula can be evaluated numerically.

Example 2.6.1 *Mean and Variance of an Inverse Gamma*

Because a gamma random variable X with intensity λ and shape parameter β has Laplace transform $L(t) = [\lambda/(t+\lambda)]^\beta$, formula (2.10) gives

$$\begin{aligned} \mathrm{E}\left(X^{-1}\right) &= \int_0^\infty \left(\frac{\lambda}{t+\lambda}\right)^\beta dt \\ &= \frac{\lambda}{\beta-1} \end{aligned}$$

for $\beta > 1$ and

$$\mathrm{E}\left(X^{-2}\right) = \int_0^\infty t\left(\frac{\lambda}{t+\lambda}\right)^\beta dt$$

$$
\begin{aligned}
&= \int_0^\infty \lambda\Big(\frac{\lambda}{t+\lambda}\Big)^{\beta-1} dt - \int_0^\infty \lambda\Big(\frac{\lambda}{t+\lambda}\Big)^{\beta} dt \\
&= \frac{\lambda^2}{\beta-2} - \frac{\lambda^2}{\beta-1}
\end{aligned}
$$

for $\beta > 2$. It follows that $\mathrm{Var}(X^{-1}) = \lambda^2/[(\beta-1)^2(\beta-2)]$ for $\beta > 2$. ■

To calculate the expectation of a ratio X^m/Y^n for a positive random variable Y and an arbitrary random variable X, we consider the mixed characteristic function and Laplace transform $M(s,t) = \mathrm{E}(e^{isX-tY})$. Assuming that $\mathrm{E}(|X|^m) < \infty$ for some positive integer m, we can write

$$
\frac{\partial^m}{\partial s^m} M(s,t) = \mathrm{E}\left[(iX)^m e^{isX-tY}\right]
$$

by virtue of Example 1.2.5 with dominating random variable $|X|^k e^{-tY}$ for the kth partial derivative. For $n > 0$ and $\mathrm{E}(|X|^m Y^{-n}) < \infty$, we now invoke Fubini's theorem and calculate

$$
\begin{aligned}
\int_0^\infty t^{n-1} \frac{\partial^m}{\partial s^m} M(0,t)\, dt &= \int_0^\infty t^{n-1}\, \mathrm{E}\left[(iX)^m e^{-tY}\right] dt \\
&= \mathrm{E}\left[\int_0^\infty t^{n-1}(iX)^m e^{-tY}\, dt\right] \\
&= \mathrm{E}\left[\frac{(iX)^m}{Y^n}\int_0^\infty r^{n-1}e^{-r}\, dr\right] \\
&= \mathrm{E}\left[\frac{(iX)^m}{Y^n}\right]\Gamma(n).
\end{aligned}
$$

Rearranging this yields

$$
\mathrm{E}\left(\frac{X^m}{Y^n}\right) = \frac{1}{i^m\Gamma(n)}\int_0^\infty t^{n-1}\frac{\partial^m}{\partial s^m}M(0,t)\, dt. \tag{2.11}
$$

Example 2.6.2 *Mean of a Beta Random Variable*

If U and V are independent gamma random variables with common intensity λ and shape parameters α and β, then the ratio $U/(U+V)$ has a beta distribution with parameters α and β. The reader is asked to prove this fact in Problem 32. It follows that the mixed characteristic function and Laplace transform

$$
\begin{aligned}
M_{U,U+V}(s,t) &= \mathrm{E}\left[e^{isU-t(U+V)}\right] \\
&= \mathrm{E}\left[e^{(is-t)U}\right]\mathrm{E}\left[e^{-tV}\right] \\
&= \Big(\frac{\lambda}{\lambda-is+t}\Big)^{\alpha}\Big(\frac{\lambda}{\lambda+t}\Big)^{\beta}.
\end{aligned}
$$

Equation (2.11) consequently gives the mean of the beta distribution as

$$\begin{aligned}
\mathrm{E}\left(\frac{U}{U+V}\right) &= \frac{1}{i}\int_0^\infty \alpha i \frac{\lambda^\alpha}{(\lambda+t)^{\alpha+1}}\left(\frac{\lambda}{\lambda+t}\right)^\beta dt \\
&= -\frac{\alpha\lambda^{\alpha+\beta}}{(\alpha+\beta)(\lambda+t)^{\alpha+\beta}}\Big|_0^\infty \\
&= \frac{\alpha}{\alpha+\beta}.
\end{aligned}$$

2.7 Reduction of Degree

The method of reduction of degree also involves recurrence relations. However, instead of creating these by conditioning, we now employ integration by parts and simple algebraic transformations. The Stein and Chen lemmas given below find their most important applications in the approximation theories featured in the books [18, 187].

Example 2.7.1 *Stein's Lemma*

Suppose X is normally distributed with mean μ and variance σ^2 and $g(x)$ is a differentiable function such that $|g(X)(X-\mu)|$ and $|g'(X)|$ have finite expectations. Stein's lemma [187] asserts that

$$\mathrm{E}[g(X)(X-\mu)] = \sigma^2\,\mathrm{E}[g'(X)].$$

To prove this formula, we note that integration by parts produces

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^\infty g(x)(x-\mu)e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \\
&= \lim_{a_n\to-\infty}\lim_{b_n\to\infty}\left[\frac{-\sigma^2 g(x)e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\Bigg|_{a_n}^{b_n} + \frac{\sigma^2}{\sqrt{2\pi\sigma^2}}\int_{a_n}^{b_n} g'(x)e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\right] \\
&= \frac{\sigma^2}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^\infty g'(x)e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx.
\end{aligned}$$

The boundary terms vanish for carefully chosen sequences a_n and b_n tending to $\pm\infty$ because the integrable function $|g(x)(x-\mu)|\exp[-\frac{(x-\mu)^2}{2\sigma^2}]$ cannot be bounded away from 0 as $|x|$ tends to ∞. To illustrate the repeated application of Stein's lemma, take $g(x) = (x-\mu)^{2n-1}$. Then the important moment identity

$$\begin{aligned}
\mathrm{E}[(X-\mu)^{2n}] &= \sigma^2(2n-1)\,\mathrm{E}[(X-\mu)^{2n-2}] \\
&= \sigma^{2n}(2n-1)(2n-3)\cdots 1
\end{aligned}$$

follows immediately. ■

Example 2.7.2 *Reduction of Degree for the Gamma*

A random variable X with gamma density $\lambda^\alpha x^{\alpha-1}e^{-\lambda x}/\Gamma(\alpha)$ on $(0,\infty)$ satisfies the analogous reduction of degree formula

$$\mathrm{E}[g(X)X] = \frac{1}{\lambda}\,\mathrm{E}[g'(X)X] + \frac{\alpha}{\lambda}\,\mathrm{E}[g(X)].$$

Provided the required moments exist and $\lim_{x\to 0} g(x)x^\alpha = 0$, the integration by parts calculation

$$\begin{aligned}
&\frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty g(x)xx^{\alpha-1}e^{-\lambda x}\,dx \\
= {} & \frac{\lambda^\alpha}{\Gamma(\alpha)\lambda}\Big[-g(x)xx^{\alpha-1}e^{-\lambda x}\Big|_0^\infty + \int_0^\infty g'(x)xx^{\alpha-1}e^{-\lambda x}\,dx \\
&+ \int_0^\infty g(x)\alpha x^{\alpha-1}e^{-\lambda x}\,dx\Big]
\end{aligned}$$

is valid. The special case $g(x) = x^{n-1}$ yields the recurrence relation

$$\mathrm{E}(X^n) = \frac{(n-1+\alpha)}{\lambda}\,\mathrm{E}(X^{n-1})$$

for the moments of X. ■

Example 2.7.3 *Chen's Lemma*

Chen [36] pursues the formula $\mathrm{E}[Zg(Z)] = \lambda\,\mathrm{E}[g(Z+1)]$ for a Poisson random variable Z with mean λ as a kind of discrete analog to Stein's lemma. The proof of Chen's result

$$\begin{aligned}
\sum_{j=0}^\infty jg(j)\frac{\lambda^j e^{-\lambda}}{j!} &= \lambda\sum_{j=1}^\infty g(j)\frac{\lambda^{j-1}e^{-\lambda}}{(j-1)!} \\
&= \lambda\sum_{k=0}^\infty g(k+1)\frac{\lambda^k e^{-\lambda}}{k!}
\end{aligned}$$

is almost trivial. The choice $g(z) = z^{n-1}$ gives the recurrence relation

$$\begin{aligned}
\mathrm{E}(Z^n) &= \lambda\,\mathrm{E}[(Z+1)^{n-1}] \\
&= \lambda\sum_{k=0}^{n-1}\binom{n-1}{k}\mathrm{E}(Z^k)
\end{aligned}$$

for the moments of Z. ■

2.8 Spherical Surface Measure

In this section and the next, we explore probability measures on surfaces. Surface measures are usually treated using differential forms and manifolds [185]. With enough symmetry, one can dispense with these complicated mathematical objects and fall back on integration on R^n. This concrete approach has the added benefit of facilitating the calculation of certain expectations.

Let $g(\|x\|)$ be any probability density such as $e^{-\pi\|x\|^2}$ on R^n that depends only on the Euclidean distance $\|x\|$ of a point x from the origin. Given a choice of $g(\|x\|)$, one can define the integral of a continuous, real-valued function $f(s)$ on the unit sphere $S_{n-1} = \{x \in \mathsf{R}^n : \|x\| = 1\}$ by

$$\int_{S_{n-1}} f(s)\, d\omega_{n-1}(s) \quad = \quad a_{n-1} \int f\left(\frac{x}{\|x\|}\right) g(\|x\|)\, dx \tag{2.12}$$

for a positive constant a_{n-1} to be specified [16]. It is trivial to show that this yields an invariant integral in the sense that

$$\int_{S_{n-1}} f(Ts)\, d\omega_{n-1}(s) \quad = \quad \int_{S_{n-1}} f(s)\, d\omega_{n-1}(s)$$

for any orthogonal transformation T. In this regard note that $|\det(T)| = 1$ and $\|Tx\| = \|x\|$. Taking $f(s) = 1$ produces a total mass of a_{n-1} for the surface measure ω_{n-1}.

Of course, the constant a_{n-1} is hardly arbitrary. We can pin it down by proving the product measure formula

$$\int h(x)\, dx \quad = \quad \int_0^\infty \int_{S_{n-1}} h(rs)\, d\omega_{n-1}(s)\, r^{n-1}\, dr \tag{2.13}$$

for any integrable function $h(x)$ on R^n. Formula (2.13) says that we can integrate over R^n by cumulating the surface integrals over successive spherical shells. To prove (2.13), we interchange orders of integration as needed and execute the successive changes of variables $t = r\|x\|^{-1}$, $z = tx$, and $t = \|z\|r^{-1}$. These maneuvers turn the right-hand side of formula (2.13) into

$$\begin{aligned}
& \int_0^\infty \int_{S_{n-1}} h(rs)\, d\omega_{n-1}(s)\, r^{n-1}\, dr \\
= \quad & a_{n-1} \int_0^\infty \int h(rx/\|x\|) g(\|x\|)\, dx\, r^{n-1}\, dr \\
= \quad & a_{n-1} \int \int_0^\infty h(rx/\|x\|)\, r^{n-1}\, dr\, g(\|x\|)\, dx \\
= \quad & a_{n-1} \int \int_0^\infty h(tx)(t\|x\|)^{n-1} \|x\|\, dt\, g(\|x\|)\, dx
\end{aligned}$$

$$\begin{aligned}
&= a_{n-1}\int_0^\infty \int h(tx)\|x\|^n g(\|x\|)\,dx\,t^{n-1}\,dt \\
&= a_{n-1}\int_0^\infty \int h(z)(\|z\|/t)^n g(\|z\|/t)\,t^{-n}\,dz\,t^{n-1}\,dt \\
&= a_{n-1}\int \int_0^\infty (\|z\|/t)^{n-1} g(\|z\|/t)\|z\|t^{-2}\,dt\,h(z)\,dz \\
&= a_{n-1}\int_0^\infty r^{n-1}g(r)\,dr \int h(z)\,dz.
\end{aligned}$$

This establishes equality (2.13) provided we take $a_{n-1}\int_0^\infty r^{n-1}g(r)\,dr = 1$. For the choice $g(r) = e^{-\pi r^2}$, we calculate

$$\int_0^\infty r^{n-1}e^{-\pi r^2}dr \;=\; \int_0^\infty \left(\frac{t}{\pi}\right)^{(n-2)/2} e^{-t}\frac{1}{2\pi}\,dt \;=\; \frac{\Gamma(\frac{n}{2})}{2\pi^{n/2}}. \tag{2.14}$$

Thus, the surface area a_{n-1} of S_{n-1} reduces to $2\pi^{n/2}/\Gamma(\frac{n}{2})$. Omitting the constant a_{n-1} in the definition (2.12) yields the uniform probability distribution on S_{n-1}.

Besides offering a method of evaluating integrals, formula (2.13) demonstrates that the definition of surface measure does not depend on the choice of the function $g(\|x\|)$. In fact, consider the extension

$$h(x) \;=\; f\left(\frac{x}{\|x\|}\right)1_{\{1\le\|x\|\le c\}}$$

of a function $f(x)$ on S_{n-1}. If we take $c = \sqrt[n]{n+1}$, then $\int_1^c r^{n-1}\,dr = 1$, and formula (2.13) amounts to

$$\int_{S_{n-1}} f(s)\,d\omega_{n-1}(s) \;=\; \frac{\int h(x)\,dx}{\int_1^c r^{n-1}dr} \;=\; \int h(x)\,dx,$$

which, as Baker notes [16], affords a definition of the surface integral that does not depend on the choice of the probability density $g(\|x\|)$. As a by-product of this result, it follows that the surface area a_{n-1} of S_{n-1} also does not depend on $g(\|x\|)$.

Example 2.8.1 *Moments of* $\|x\|$ *Relative to* $e^{-\pi\|x\|^2}$

Formula (2.13) gives

$$\begin{aligned}
\int \|x\|^k e^{-\pi\|x\|^2}dx &= \int_0^\infty \int_{S_{n-1}} r^k e^{-\pi r^2}\,d\omega_{n-1}(s) r^{n-1}dr \\
&= a_{n-1}\int_0^\infty r^{n+k-1}e^{-\pi r^2}dr \\
&= \frac{a_{n-1}}{a_{n+k-1}}.
\end{aligned}$$

Negative as well as positive values of $k > -n$ are permitted. ■

Example 2.8.2 *Integral of a Polynomial*

The function $f(x) = x_1^{k_1} \cdots x_n^{k_n}$ is a monomial when $k_1, \ldots, k_n$ are non-negative integers. A linear combination of monomials is a polynomial. To find the integral of $f(x)$ on S_{n-1}, it is convenient to put $k = \sum_{j=1}^n k_j$ and use the probability density $g(\|x\|) = a_{n+k-1}\|x\|^k e^{-\pi\|x\|^2}/a_{n-1}$. With these choices,

$$
\begin{aligned}
\int_{S_{n-1}} f(s)\, d\omega_{n-1}(s) &= a_{n-1}\frac{a_{n+k-1}}{a_{n-1}} \int \frac{x_1^{k_1}\cdots x_n^{k_n}}{\|x\|^k}\|x\|^k e^{-\pi\|x\|^2}\, dx \\
&= a_{n+k-1} \int x_1^{k_1}\cdots x_n^{k_n} e^{-\pi\|x\|^2}\, dx \\
&= a_{n+k-1} \prod_{j=1}^n \int_{-\infty}^{\infty} x_j^{k_j} e^{-\pi x_j^2}\, dx_j .
\end{aligned}
$$

If any k_j is odd, then the corresponding one-dimensional integral in the last product vanishes. Hence, the surface integral of the monomial vanishes as well. If all k_j are even, then the same reasoning that produced equation (2.14) leads to

$$
\begin{aligned}
\int_{-\infty}^{\infty} x_j^{k_j} e^{-\pi x_j^2}\, dx_j &= 2\int_0^{\infty} x_j^{k_j} e^{-\pi x_j^2}\, dx_j \\
&= \frac{\Gamma(\frac{k_j+1}{2})}{\pi^{(k_j+1)/2}}.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\int_{S_{n-1}} x_1^{k_1}\cdots x_n^{k_n}\, d\omega_{n-1}(s) &= \frac{2\pi^{(n+k)/2}}{\Gamma(\frac{n+k}{2})} \prod_{j=1}^n \frac{\Gamma(\frac{k_j+1}{2})}{\pi^{(k_j+1)/2}} \\
&= \frac{2\prod_{j=1}^n \Gamma(\frac{k_j+1}{2})}{\Gamma(\frac{n+k}{2})}
\end{aligned}
$$

when all k_j are even. ■

2.9 Dirichlet Distribution

The Dirichlet distribution generalizes the beta distribution. As such, it lives on the unit simplex $T_{n-1} = \{x \in \mathsf{R}_+^n : \|x\|_1 = 1\}$, where $\|x\|_1 = \sum_{j=1}^n |x_j|$ and

$$
\mathsf{R}_+^n = \{x \in \mathsf{R}^n : x_j > 0,\ j = 1, \ldots, n\}.
$$

By analogy with our definition (2.12) of spherical surface measure, one can define the simplex surface measure μ_{n-1} on T_{n-1} through the equation

$$\int_{T_{n-1}} f(s)\, d\mu_{n-1}(s) \quad = \quad b_{n-1} \int_{\mathsf{R}_+^n} f\left(\frac{x}{\|x\|_1}\right) g(\|x\|_1)\, dx \qquad (2.15)$$

for any continuous function $f(s)$ on T_{n-1}. In this setting, $g(\|x\|_1)$ is a probability density on R_+^n that depends only on the distance $\|x\|_1$ of x from the origin.

One can easily show that this definition of surface measure is invariant under permutation of the coordinates. One can also prove the product measure formula

$$\int_{\mathsf{R}_+^n} h(x)\, dx \quad = \quad \frac{1}{\sqrt{n}} \int_0^\infty \int_{T_{n-1}} h(rs)\, d\mu_{n-1}(s)\, r^{n-1}\, dr. \qquad (2.16)$$

The appearance of the factor $1/\sqrt{n}$ here can be explained by appealing to geometric intuition. In formula (2.16) we integrate $h(x)$ by summing its integrals over successive slabs multiplied by the thicknesses of the slabs. Now the thickness of a slab amounts to nothing more than the distance between two slices $(r + dr)T_{n-1}$ and rT_{n-1}. Given that the corresponding centers of mass are $(r + dr)n^{-1}\mathbf{1}$ and $rn^{-1}\mathbf{1}$, the slab thickness is $dr/\sqrt{n}$.

The proof of formula (2.16) is virtually identical to the proof of formula (2.13). In the final stage of the proof, we must set

$$\frac{b_{n-1}}{\sqrt{n}} \int_0^\infty r^{n-1} g(r)\, dr \quad = \quad 1.$$

The choice $g(r) = e^{-r}$ immediately gives $\int_0^\infty r^{n-1}e^{-r}dr = \Gamma(n)$. It follows that the surface area b_{n-1} of T_{n-1} is $\sqrt{n}/\Gamma(n)$. Omitting the constant b_{n-1} in the definition (2.15) yields the uniform probability distribution on T_{n-1}.

As before we evaluate the moment

$$\begin{aligned}
\int_{\mathsf{R}_+^n} \|x\|_1^k e^{-\|x\|_1} dx \quad &= \quad \frac{1}{\sqrt{n}} \int_0^\infty \int_{T_{n-1}} r^k e^{-r} d\mu_{n-1}(s) r^{n-1} dr \\
&= \quad \frac{b_{n-1}}{\sqrt{n}} \int_0^\infty r^{n+k-1} e^{-r} dr \\
&= \quad \frac{\Gamma(n+k)}{\Gamma(n)}.
\end{aligned}$$

For the multinomial $f(x) = x_1^{k_1} \cdots x_n^{k_n}$ with $k = \sum_{j=1}^n k_j$, we then evaluate

$$\begin{aligned}
\int_{T_{n-1}} f(s)\, d\mu_{n-1}(s) \quad &= \quad b_{n-1} \frac{\Gamma(n)}{\Gamma(n+k)} \int_{\mathsf{R}_+^n} \frac{x_1^{k_1} \cdots x_n^{k_n}}{\|x\|_1^k} \|x\|_1^k e^{-\|x\|_1} dx \\
&= \quad b_{n-1} \frac{\Gamma(n)}{\Gamma(n+k)} \int_{\mathsf{R}_+^n} x_1^{k_1} \cdots x_n^{k_n} e^{-\|x\|_1} dx
\end{aligned}$$

$$= \quad b_{n-1}\frac{\Gamma(n)}{\Gamma(n+k)}\prod_{j=1}^{n}\Gamma(k_j+1)$$

using the probability density

$$g(\|x\|_1) \quad = \quad \frac{\Gamma(n)}{\Gamma(n+k)}\|x\|_1^k e^{-\|x\|_1}$$

on R_+^n. This calculation identifies the Dirichlet distribution

$$\frac{\Gamma(k)}{b_{n-1}\Gamma(n)\prod_{j=1}^n \Gamma(k_j)}\prod_{j=1}^{n} s_j^{k_j-1}$$

as a probability density on T_{n-1} relative to μ_{n-1} with moments

$$\mathrm{E}\left(s_1^{l_1}\cdots s_n^{l_n}\right) \quad = \quad \frac{\Gamma(k)\prod_{j=1}^n \Gamma(k_j+l_j)}{\Gamma(k+l)\prod_{j=1}^n \Gamma(k_j)},$$

where $l = \sum_{j=1}^n l_j$. Note that $k_j > 0$ need not be an integer.

2.10 Problems

1. Let X represent the number of fixed points of a random permutation of the set $\{1,\ldots,n\}$. Demonstrate that X has the falling factorial moment

$$\mathrm{E}[X(X-1)\cdots(X-k+1)] \quad = \quad k!\,\mathrm{E}\left[\binom{X}{k}\right] \quad = \quad 1$$

for $1 \le k \le n$. (Hints: Note that $\binom{X}{k}$ is the number of ways of choosing k points among the available fixed points. Choose the points first, and then calculate the probability that they are fixed.)

2. In a certain building, p people enter an elevator stationed on the ground floor. There are n floors above the ground floor, and each is an equally likely destination. If the people exit the elevator independently, then show that the elevator makes on average

$$n\Big[1-\Big(1-\frac{1}{n}\Big)^p\Big]$$

stops in discharging all p people.

3. Numbers are drawn randomly from the set $\{1,2,\ldots,n\}$ until their sum exceeds k for $0 \le k \le n$. Show that the expected number of draws equals

$$e_k \quad = \quad \Big(1+\frac{1}{n}\Big)^k.$$

In particular, $e_n \approx e$. (Hint: Show that $e_k = 1 + \frac{1}{n}[e_0 + \cdots + e_{k-1}]$.)

4. Show that the beta-binomial distribution of Example 2.3.1 has discrete density

$$\Pr(S_n = k) = \binom{n}{k} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+k)\Gamma(\beta+n-k)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha+\beta+n)}.$$

5. Prove that the $\ln X_n = \sum_{i=1}^n \ln U_i$ random variable of Example 2.3.2 follows a $-\frac{1}{2}\chi^2_{2n}$ distribution.

6. A noncentral chi-square random variable X has a χ^2_{n+2Y} distribution conditional on a Poisson random variable Y with mean λ. Show that $\mathrm{E}(X) = n + 2\lambda$ and $\mathrm{Var}(X) = 2n + 8\lambda$.

7. Consider an urn with $b \geq 1$ black balls and $w \geq 0$ white balls. Balls are extracted from the urn without replacement until a black ball is encountered. Show that the number of balls N_{bw} extracted has mean $\mathrm{E}(N_{bw}) = (b+w+1)/(b+1)$. (Hint: Derive a recurrence relation and boundary conditions for $\mathrm{E}(N_{bw})$ and solve.)

8. Give a recursive method for computing the second moments $\mathrm{E}(N_{sd}^2)$ in the family planning model.

9. In the family planning model, suppose the couple has an upper limit m on the number of children they can afford. Hence, they stop whenever they reach their goal of s sons and d daughters or m total children, whichever comes first. Let N_{sdm} now be their random number of children. Give a recursive method for computing $\mathrm{E}(N_{sdm})$.

10. In table tennis suppose that player B wins a set with probability p and player A wins a set with probability $q = 1 - p$. Each set counts 1 point. The winner of a match is the first to accumulate 21 points and at least 2 points more than the opposing player. How can one calculate the probability that player B wins? Assume that A has already accumulated i points and B has already accumulated j points. Let w_{ij} denote the probability that B wins the match. Let t_{ij} denote the corresponding expected number of further points scored before the match ends. Show that these quantities satisfy the recurrences

$$\begin{aligned} w_{ij} &= pw_{i,j+1} + qw_{i+1,j} \\ t_{ij} &= 1 + pt_{i,j+1} + qt_{i+1,j} \end{aligned}$$

for i and j between 0 and 20. The first recurrence allows one to compute w_{00} from the boundary values $w_{i,21}$ and $w_{21,j}$, where $i \leq 21$ and $j \leq 21$. In these situations, the quota of 21 total points is irrelevant, and only the excess points criterion is operative. The second recurrence has similar implications for the t_{ij}. On the boundary, the winning probability reduces to either 0 or 1 or to the hitting probability considered in Problem 38 of Chapter 7. Using the results stated there, tabulate or graph w_{00} and t_{00} as a function of p.

11. Consider the integral

$$I(a,p,y) = \int_{-\infty}^{y} \frac{1}{(a+x^2)^p}\,dx$$

for $p > \frac{1}{2}$ and $a > 0$. As an example of the method of parametric differentiation [29], prove that

$$I(a,p+n,y) = \frac{(-1)^n}{p(p+1)\cdots(p+n-1)}\frac{d^n}{da^n}I(a,p,y).$$

In the particular case $p = \frac{3}{2}$, show that

$$I\Big(a,\frac{3}{2},y\Big) = \frac{y}{a\sqrt{a+y^2}} + \frac{1}{a}.$$

Use these facts to show that the t-distribution with $2m$ degrees of freedom has finite expansion

$$\begin{aligned} &\frac{\Gamma(m+1/2)}{\sqrt{2\pi m}\Gamma(m)}\int_{-\infty}^{y}\Big(1+\frac{x^2}{2m}\Big)^{-m-1/2}dx \\ &= \frac{1}{2\sqrt{2m}}\left[\frac{y}{\sqrt{\pi}}\sum_{j=0}^{m-1}\frac{\Gamma(j+1/2)}{j!}\Big(1+\frac{y^2}{2m}\Big)^{-j-1/2} + \sqrt{2m}\right]. \end{aligned}$$

(Hints: For the case $p = \frac{3}{2}$ apply the fundamental theorem of calculus. To expand the t-distribution, use Leibniz's rule for differentiating a product.)

12. Let X be a nonnegative integer-valued random variable with probability generating function $Q(s)$. Find the probability generating functions of $X+k$ and kX in terms of $Q(s)$ for any nonnegative integer k.

13. Let $S_n = X_1 + \cdots + X_n$ be the sum of n independent random variables, each distributed uniformly over the set $\{1, 2, \ldots, m\}$. Find the probability generating function of S_n, and use it to calculate $\mathrm{E}(S_n)$ and $\mathrm{Var}(S_n)$.

14. Let $X_1, X_2, \ldots$ be an i.i.d. sequence of Bernoulli random variables with success probability p. Thus, $X_i = 1$ with probability p, and $X_i = 0$ with probability $1-p$. Demonstrate that the infinite series $S = \sum_{i=1}^{\infty} 2^{-i}X_i$ has mean p and variance $\frac{1}{3}p(1-p)$. When $p = \frac{1}{2}$,

the random variable S is uniformly distributed on $[0,1]$, and X_i can be interpreted as the ith binary digit of S [192]. In this special case also prove the well-known identity

$$\frac{\sin\theta}{\theta} = \prod_{j=1}^{\infty}\cos\left(\frac{\theta}{2^j}\right)$$

by calculating the characteristic function of $S-\frac{1}{2}$ in two different ways.

15. Consider a sequence $X_1, X_2, \ldots$ of independent, integer-valued random variables with common logarithmic distribution

$$\Pr(X_i = k) = -\frac{q^k}{k\ln(1-q)}$$

for $k \geq 1$. Let N be a Poisson random variable with mean λ that is independent of the X_i. Show that the random sum $S_N = \sum_{i=1}^{N} X_i$ has a negative binomial distribution that counts only failures. Note that the required "number of successes" in the negative binomial need not be an integer [59].

16. Suppose X has gamma density $\lambda^\beta x^{\beta-1}e^{-\lambda x}/\Gamma(\beta)$ on $(0,\infty)$, where β is not necessarily an integer. Show that X has characteristic function $(\frac{\lambda}{\lambda-it})^\beta$ and Laplace transform $(\frac{\lambda}{\lambda+t})^\beta$. Use either of these to calculate the mean and variance of X. (Hint: For the characteristic function, derive and solve a differential equation. Alternatively, calculate the Laplace transform directly by integration and show that it can be extended to an analytic function in a certain region of the complex plane.)

17. Let X have the gamma density defined in Problem 16. Conditional on X, let Y have a Poisson distribution with mean X. Prove that Y has probability generating function

$$\mathrm{E}(s^Y) = \left(\frac{\lambda}{\lambda+1-s}\right)^\beta.$$

18. Show that the bilateral exponential density $\frac{1}{2}e^{-|x|}$ has characteristic function $1/(1+t^2)$. Use this fact to calculate its mean and variance.

19. Example 2.4.6 shows that it is impossible to write a random variable U uniformly distributed on $[-1,1]$ as the difference of two i.i.d. random variables X and Y. It is also true that it is impossible to write U as the sum of two i.i.d. random variables X and Y. First of all it is clear that X and Y have support on $[-1/2,1/2]$. Hence, they possess

moments of all orders, and it is possible to represent the characteristic function of X by the series

$$\mathrm{E}\left(e^{itX}\right) = \sum_{n=0}^{\infty} \mathrm{E}(X^n)\frac{(it)^n}{n!}.$$

If one can demonstrate that the odd moments of X vanish, then it follows that its characteristic function is real and that

$$\mathrm{E}\left(e^{itU}\right) = \left[\mathrm{E}\left(e^{itX}\right)\right]^2$$

can never be negative. This contradicts the fact that $t^{-1}\sin t$ oscillates in sign. Supply the missing steps in this argument. (Hints: Why does $\mathrm{E}(X) = 0$? Assuming this is true, take n odd, expand $\mathrm{E}[(X+Y)^n]$ by the binomial theorem, and apply induction.)

20. Calculate the Laplace transform of the probability density

$$\frac{1+a^2}{a^2}e^{-x}[1-\cos(ax)]1_{\{x\geq 0\}}.$$

21. Card matching is one way of testing extrasensory perception (ESP). The tester shuffles a deck of cards labeled 1 through n and turns cards up one by one. The subject is asked to guess the value of each card and is told whenever he or she gets a match. No information is revealed for a nonmatch. According to Persi Diaconis, the optimal strategy the subject can adopt is to guess the value 1 until it turns up, then guess the value 2 until it turns up, then guess the value 3 until it turns up, and so forth. Note that this strategy gives a single match if card 2 is turned up before card 1. Show that the first two moments of the number of matches X are

$$\begin{aligned}\mathrm{E}(X) &= \sum_{j=1}^{n}\frac{1}{j!} \approx e-1 \\ \mathrm{E}(X^2) &= 2\sum_{j=1}^{n}\frac{1}{(j-1)!} - \mathrm{E}(X) \approx e+1.\end{aligned}$$

(Hint: Why does the tail probability $\Pr(X \geq j)$ equal $\frac{1}{j!}$?)

22. Suppose the right-tail probability of a nonnegative random variable X satisfies $|1-F(x)| \leq cx^{-n-\epsilon}$ for all sufficiently large x, where n is a positive integer, and ϵ and c are positive real numbers. Show that $\mathrm{E}(X^n)$ is finite.

23. Let the positive random variable X have Laplace transform $L(t)$. Prove that $\mathrm{E}[(aX+b)^{-1}] = \int_0^\infty e^{-bt}L(at)\,dt$ for $a \geq 0$ and $b > 0$.

24. Let X be a nonnegative integer-valued random variable with probability generating function $G(u)$. Prove that

$$\begin{aligned} & \mathrm{E}\left[\frac{1}{(X+k+j)(X+k+j-1)\cdots(X+k)}\right] \\ = & \frac{1}{j!}\int_0^1 u^{k-1}(1-u)^j G(u)\,du \end{aligned}$$

by taking the expectation of $\int_0^1 u^{X+k-1}(1-u)^j\,du$.

25. Suppose X has a binomial distribution with success probability p over n trials. Show that

$$\mathrm{E}\left(\frac{1}{X+1}\right) = \frac{1-(1-p)^{n+1}}{(n+1)p}.$$

26. Let χ^2_n and χ^2_{n+2} be chi-square random variables with n and $n+2$ degrees of freedom, respectively. Demonstrate that

$$\mathrm{E}[f(\chi^2_n)] = n\,\mathrm{E}\left[\frac{f(\chi^2_{n+2})}{\chi^2_{n+2}}\right]$$

for any well-behaved function $f(x)$ for which the two expectations exist. Use this identity to calculate the mean and variance of χ^2_n [34].

27. Suppose X has a binomial distribution with success probability p over n trials. Show that

$$\mathrm{E}[Xf(X)] = \frac{p}{1-p}\,\mathrm{E}\Big[(n-X)f(X+1)\Big]$$

for any function $f(x)$. Use this identity to calculate the mean and variance of X.

28. Consider a negative binomial random variable X with density

$$\Pr(X=k) = \binom{k-1}{n-1}p^n q^{k-n}$$

for $q = 1-p$ and $k \geq n$. Prove that for any function $f(x)$

$$\mathrm{E}[qf(X)] = \mathrm{E}\left[\frac{(X-n)f(X-1)}{X-1}\right].$$

Use this identity to calculate the mean of X [100].

29. Demonstrate that the unit ball $\{x \in \mathsf{R}^n : \|x\| \leq 1\}$ has volume $\pi^{n/2}/\Gamma(n/2+1)$ and the standard simplex $\{x \in \mathsf{R}^n_+ : \|x\|_1 \leq 1\}$ has volume $1/n!$.

30. An Epanechnikov random vector X has density

$$f(x) = \frac{n+2}{2v_n}\left(1 - \|x\|^2\right)$$

supported on the unit ball $\{x \in \mathsf{R}^n : \|x\| \leq 1\}$. Here v_n is the volume of the ball as given in Problem 29. Demonstrate that X has mean vector $\mathbf{0}$ and variance matrix $\frac{1}{n+4}I$, where I is the identity matrix. (Hint: Use equation (2.13) and Example 2.8.2.)

31. Suppose the random vector X is uniformly distributed on the unit simplex T_{n-1}. Let m be a positive integer with $m < n$ and v be a vector in R^n with positive components. Show that the expected value of $(v^tX)^{-m}$ is

$$\mathrm{E}[(v^tX)^{-m}] = m\binom{n-1}{m}\int_0^\infty t^{m-1}\prod_{i=1}^n \frac{1}{tv_i+1}\,dt.$$

See the article [158] for explicit evaluation of the last one-dimensional integral. (Hints: Show that

$$\mathrm{E}[(v^tX)^{-m}] = \frac{\Gamma(n)}{\Gamma(n-m)}\int_{\mathsf{R}_+^n}(v^tx)^{-m}\|x\|_1^m\frac{1}{\|x\|_1^m}e^{-\|x\|_1}\,dx.$$

Then let $s_{n-1} = \sum_{i=2}^n v_ix_i$, and demonstrate that

$$\int_0^\infty \frac{1}{(v_1x_1+s_{n-1})^m}e^{-x_1}\,dx_1 = \frac{1}{\Gamma(m)}\int_0^\infty \frac{t^{m-1}}{tv_1+1}e^{-ts_{n-1}}\,dt$$

by invoking equation (2.10).)

32. One can generate the Dirichlet distribution by a different mechanism than the one developed in the text [114]. Take n independent gamma random variables $X_1, \ldots, X_n$ of unit scale and form the ratios

$$Y_i = \frac{X_i}{\sum_{j=1}^n X_j}.$$

Here X_i has density $x_i^{k_i-1}e^{-x_i}/\Gamma(k_i)$ on $(0,\infty)$ for some $k_i > 0$. Clearly, each $Y_i \geq 0$ and $\sum_{i=1}^n Y_i = 1$. Show that $(Y_1, \ldots, Y_n)^t$ follows a Dirichlet distribution on T_{n-1}.

33. Continuing Problem 32, calculate $\mathrm{E}(Y_i)$, $\mathrm{Var}(Y_i)$, and $\mathrm{Cov}(Y_i, Y_j)$ for $i \neq j$. Also show that $(Y_1 + Y_2, Y_3, \ldots, Y_n)^t$ has a Dirichlet distribution.

34. Continuing Problem 32, prove that the random variables

$$\frac{X_1}{X_1+X_2},\ \frac{X_1+X_2}{X_1+X_2+X_3},\ \ldots,\ \frac{X_1+\cdots+X_{n-1}}{X_1+\cdots+X_n},\ X_1+\cdots+X_n$$

are independent. What distributions to these random variables follow? (Hints: Denote the random variables $Z_1, \ldots, Z_n$. Make a multidimensional change of variables to find their joint density using the identities $X_1 = \prod_{i=1}^n Z_i$ and $X_j = (1 - Z_{j-1}) \prod_{i=j}^n Z_i$ for $j > 1$.)