Part II

Integration

Applications of Integration in Statistics

The most common use of integration methods in statistics is in Bayesian analysis. For the posterior computation

$$\pi(\theta|\boldsymbol{y}) = \frac{L(\theta;\boldsymbol{y})\pi(\theta)}{\int_\Theta L(t;\boldsymbol{y})\pi(t)\,dt}$$

where the denominator is the *marginal likelihood* or *prior predictive distribution*

$$M(\boldsymbol{Y}) = \int_\Theta L(t;\boldsymbol{Y})\pi(t)\,dt$$

This integral is over a region determined by the model and the dimension of the parameter; typically this is some open subset of $\mathbb{R}^p$.

Applications of Integration in Statistics

The ratio marginal likelihoods for two models

$$\frac{M_1(\boldsymbol{y})}{M_2(\boldsymbol{y})} = \frac{\int_\Theta L_1(t;\boldsymbol{Y})\pi_1(t)\,dt}{\int_\Theta L_2(t;\boldsymbol{Y})\pi_2(t)\,dt}$$

can be used as the basis of a model selection procedure. Model 1 is preferred to Model 2 if this ratio is greater than 1.

The ratio M_1/M_2 is termed the *Bayes factor*.

Note: the marginal likelihood measures how well the model (likelihood and prior) predict the **observed** data, that is, it is a "within-sample" measure.

Applications of Integration in Statistics

Bayesian estimation typically relies on minimizing some expected posterior loss

$$\widehat{\theta} = \underset{t}{\operatorname{argmin}} \left\{ \int_{\Theta} \mathcal{L}(t, \theta) \pi(\theta | \boldsymbol{y}) \, d\theta \right\}$$

and sometimes this results in having to compute a functional of the posterior of the form

$$\mathbb{E}_{\theta|\boldsymbol{Y}}[h(\theta)] = \int_{\Theta} h(\theta) \pi(\theta | \boldsymbol{y}) \, d\theta$$

The *posterior predictive distribution* is also of interest

$$f(\boldsymbol{y}^* | \boldsymbol{y}) = \int_{\Theta} L(\boldsymbol{y}^*; \theta) \pi(\theta | \boldsymbol{y}) \, d\theta$$

Applications of Integration in Statistics

In a non-Bayesian setting, the marginal likelihood is required for the analysis of certain sorts of longitudinal data. Suppose the following data are available:

- For individual $i = 1, \ldots, N$, responses y_{ij}, $j = 1, \ldots, n_i$, are recorded over different time points
- Covariate values $\boldsymbol{x}_{ij}$ are available at baseline and on subsequent visits.

A *linear mixed model* or *linear random effects model* is commonly used

$$Y_{ij} = \boldsymbol{x}_i^\mathsf{T} \beta + \eta_i + \epsilon_{ij}$$

with, say, $\epsilon_{ij} \sim \text{Normal}(0, \sigma_\epsilon^2)$ are (possibly correlated) residual errors, and where random effects $\eta_1, \ldots, \eta_N$ are presumed to be iid and $\eta_i \sim \text{Normal}(0, \sigma_\eta^2)$.

Applications of Integration in Statistics

Here the marginal likelihood is formed by integrating out the random effects parameters. If $\theta = (\beta, \sigma_\epsilon, \sigma_\eta)$, then for individual i,

$$L_i(\theta; \boldsymbol{y}_i, \boldsymbol{x}_i) = \int f(y_{i1} \ldots, y_{in_i} | \theta, \eta_i) p(\eta_i)\, d\eta_i$$

say, and

$$L(\theta; \boldsymbol{x}, \boldsymbol{y}) = \prod_{i=1}^{N} L_i(\theta; \boldsymbol{y}_i, \boldsymbol{x}_i)$$

Inference can then proceed using ML.

Note: this is not the only method for analyzing mixed models, it is also possible to estimate and condition upon the η_i.

Analytical Approximation Methods

In many large sample situations, the posterior distribution can be well approximated by a multivariate normal distributions.

If the normal approximation is not valid as the sample is too small, perhaps an analytical quadratic approximation along the lines of the Laplace approximation may be used.

See, in particular:

- L. Tierney and J.B. Kadane, Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, **81** (1986), pp. 82-86.
- Casella and Robert, p 107.

Basic Numerical Integration

The first method typically encountered for univariate numerical integration is the *trapezium rule:* with grid $x_0, x_1, \ldots, x_n$,

$$x_j = x_0 + jh \qquad\qquad h = (x_n - x_0)/n$$

say, if $y_j = g(x_j)$, then

$$\int_{x_0}^{x_n} g(x)\,dx \doteq h\left[\frac{(y_0 + y_n)}{2} + (y_1 + \cdots + y_{n-1})\right]$$

that is

$$\int_{x_0}^{x_n} g(x)\,dx \doteq \frac{hg(x_0) + hg(x_n)}{2} + \sum_{j=1}^{n-1}(x_j - x_{j-1})g(x_j)$$

The error in this approximation is $\mathrm{O}(h^2)$

Gaussian Quadrature

For integrals with respect to probability distributions, it is natural to consider approximations of the form

$$\int_{-\infty}^{\infty} g(x)p(x)\,dx \doteq \sum_{j=0}^{k} w_j g(x_j)$$

say, for points x_j and weights w_j chosen in line with the probability distribution. In the trapezium rule, the density p is a uniform density, and the points are uniformly distributed.

This method of approximation works well when the integrand is smooth.

Approximating the integrand well is important in evaluating the integral.

A Basic Introduction to Hilbert Spaces

Let $\mathcal{H}$ be a *vector space* with *inner product* $\langle h_1, h_2 \rangle$

- for $a, b \in \mathbb{R}$, $\langle ah_1 + bh_2, h_3 \rangle = a\langle h_1, h_2 \rangle + b\langle h_1, h_3 \rangle$,
- $\langle h_1, h_2 \rangle = \langle h_2, h_1 \rangle$,
- $\langle h_1, h_1 \rangle \geq 0$, with $\langle h_1, h_1 \rangle = 0$ if and only if $h_1 \equiv 0$.

The *norm* of $h \in \mathcal{H}$ is

$$\|h\| = \langle h_1, h_1 \rangle^{1/2}$$

Elements of $\mathcal{H}$ could be vectors in $\mathbb{R}^p$ say, in which case the usual inner product is used

$$\langle h_1, h_2 \rangle = h_1^\intercal h_2 = \sum_{j=1}^{p} h_{1j} h_{2j}$$

A final requirement is for a vector space $\mathcal{H}$ to be a *Hilbert space* is that it is *complete*, that is, for a sequence $h_1, h_2, \ldots \in \mathcal{H}$, for every $\epsilon > 0$, there exist m, n such that

$$\|h_m - h_n\| < \epsilon$$

that is,

$$\lim_{n \longrightarrow \infty} h_n = h$$

exists, for $h \in \mathcal{H}$

However, elements of $\mathcal{H}$ can be **infinite dimensional**.

For example, $\mathcal{H}$ could contain real-valued functions that obey certain integrability conditions: for example, if $h \in \mathcal{H}$, with

$$h : \mathbb{R} \longrightarrow \mathbb{R}$$

we might require

$$\int h(x)^2 \, dx < \infty$$

In this case, an inner product that is often used is

$$\langle h_1, h_2 \rangle = \int h_1(x) h_2(x) \, dx$$

This Hilbert space is denoted L^2.

A fundamental result for the Hilbert space L^2 relates to approximation: essentially, functions in L^2 can be arbitrarily well approximated. That is, if g is a square-integrable continuous function defined on $\mathbb{R}$, then g can be approximated by a function h such that

$$\|g - h\| < \epsilon$$

for ϵ arbitrarily small (that is, L^2 is *dense*).

More precisely: this result relates to functions that have ***compact*** *(bounded) support.*

A sequence $\{\psi_n\}_{n\geq 1}$ of elements of Hilbert space $\mathcal{H}$ is *orthogonal* if

$$\langle \psi_m, \psi_n \rangle = 0 \quad \text{for all } m \neq n$$

and *orthonormal* if it is orthogonal and also $\|\psi_n\| = 1$ for each n.

In a Hilbert space, $\{\psi_n\}_{n\geq 1}$ is an *orthonormal basis* if for each $g \in \mathcal{H}$

$$g = \sum_n \langle g, \psi_n \rangle \psi_n$$

In this case

$$\|g\| = \sum_n |\langle g, \psi_n \rangle|^2$$

Approximation: For each $g \in \mathcal{H}$ and $\epsilon > 0$, there exists an $m_0 = m_0(g, \epsilon)$ such that

$$\left\| g - \sum_{n=1}^{m} \langle g, \psi_n \rangle \psi_n \right\| \leq \epsilon$$

for all $m \geq m_0$.

That is, the function g can be arbitrarily well approximated using a **finite** series expansion in the orthonormal basis.

All of these results are valid if the regular integral is replaced by an integral with respect to a probability measure μ.

Square-integrability becomes

$$\int h(x)^2 \, \mu(dx) < \infty$$

and the inner product becomes

$$\langle h_1, h_2 \rangle = \int h_1(x) h_2(x) \, \mu(dx)$$

If μ admits a density $p(x)$, then

$$\langle h_1, h_2 \rangle = \int h_1(x) h_2(x) p(x) \, dx$$

Orthonormal Polynomial Bases

Many probability distributions can be associated with a natural orthonormal *polynomial* basis

$$\psi_n(x) = \sum_{j=0}^{\infty} a_{nj} x^j$$

Polynomials are useful for approximating arbitrary real functions. To construct an orthonormal polynomial representation for a probability density is more complicated, but essentially reduces to choosing the coefficients a_{nj}.

Provided the density has a *moment generating function*, it can be shown that an orthonormal polynomial basis representation exists.

Orthonormal Polynomial Bases

The orthonormal polynomials can be constructed using the Gram-Schmidt procedure: let $g_1, \ldots, g_k$ be a sequence of polynomials, and

$$\Pi(g_2|g_1) = \frac{\langle g_1, g_2 \rangle}{\langle g_1, g_1 \rangle} g_1 = \left\{ \frac{\int g_1(x) g_2(x) p(x)\, dx}{\int g_1^2(x) p(x)\, dx} \right\} g_1(x)$$

Then the sequence

$$\begin{aligned} \varphi_1 &= g_1 \\ \varphi_2 &= g_2 - \Pi(g_2|\varphi_1) \\ \varphi_3 &= g_3 - \Pi(g_3|\varphi_1) - \Pi(g_3|\varphi_2) \\ &\vdots \quad \vdots \quad \vdots \end{aligned}$$

is an orthogonal sequence, and $\psi_j = \varphi_j / \|\varphi_j\|$ is the normalized version.

Orthonormal Polynomial Bases

Moments can also be used to define the orthonormal basis. If

$$\mu_k = \mathbb{E}[X^k] = \int x^k p(x)\, dx$$

Then the polynomials $\{\varphi_n\}$, defined by

$$\varphi_n(x) = \begin{vmatrix} \mu_0 & \mu_1 & \mu_2 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_{n+1} \\ \vdots & \vdots & \vdots & & \vdots \\ \mu_{n-1} & \mu_n & \mu_{n+1} & \cdots & \mu_{2n-1} \\ 1 & x & x^2 & \cdots & x^n \end{vmatrix}$$

are orthogonal with respect to p.

Orthonormal Polynomial Bases

Example (Standard Normal Density)

Suppose

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

We need to find $\{\psi_n\}$ such that

$$\int_{-\infty}^{\infty} \psi_m(x)\psi_n(x) \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 0$$

for $m \neq n$.

Orthonormal Polynomial Bases

Example (Standard Normal Density)

Consider the exponential generating function for the $\{\psi_n\}_{n\geq 0}$

$$E(x,t) = \sum_{n=0}^{\infty} \frac{\psi_n(x)}{n!} t^n$$

so that

$$\psi_n(x) = \left. \frac{d^n E(x,t)}{dt^n} \right|_{t=0}$$

Suppose that we let

$$E(x,t) = \exp\{xt - t^2/2\}$$

in order to specify the ψ_n.

Orthonormal Polynomial Bases

Example (Standard Normal Density)

Then it follows that

$$\begin{aligned}
\int E(x,s)E(x,t)p(x)\,dx &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{xs-s^2/2}e^{xt-t^2/2}e^{-x^2/2}\,dx \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{st}e^{-(x-s-t)^2/2}\,dx \\
&= e^{st} \\
&= \sum_{n=0}^{\infty}\frac{s^n t^n}{n!}
\end{aligned}$$

Orthonormal Polynomial Bases

Example (Standard Normal Density)

However, the integral on the left-hand side can be written

$$\int \left\{ \sum_{m=0}^{\infty} \frac{\psi_m(x)}{m!} s^m \right\} \left\{ \sum_{n=0}^{\infty} \frac{\psi_n(x)}{n!} t^n \right\} p(x)\, dx$$

which can be rewritten

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left\{ \int \frac{\psi_m(x)}{m!} \frac{\psi_n(x)}{n!} p(x)\, dx \right\} s^m t^n$$

Orthonormal Polynomial Bases

Example (Standard Normal Density)

Equating the two expansions yields that

$$\int \frac{\psi_m(x)}{m!}\frac{\psi_n(x)}{n!}p(x)\,dx = \begin{cases} 0 & m \neq n \\ \dfrac{1}{n!} & m = n \end{cases}$$

and hence that the sequence $\{H_n(x)\}$

$$H_n(x) = \frac{1}{\sqrt{n!}}\psi_n(x)$$

is an orthonormal sequence with respect to this $p(x)$.

$H_n(x)$ is termed the *Hermite polynomial of order n*.

Orthonormal Polynomial Bases

Example (Standard Normal Density)

An explicit form for $H_n(x)$ can be expanding the generating function $E(x,t)$ as

$$H_n(x) = \sum_{j=0}^{\lfloor n/2 \rfloor} \frac{n!(-1)^j x^{n-2j}}{2^j j!(n-2j)!}$$

Recursion formulae can also be derived

$$H_{n+1}(x) = xH_n(x) - nH_{n-1}(x)$$

which is useful for computation.

Orthonormal Polynomial Bases

Example (Gamma Density)

If

$$p(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x} \qquad x > 0$$

then define

$$E(x,t) = \frac{1}{(1-t)^{\alpha}} \exp\left\{-\frac{tx}{1-t}\right\} = \sum_{n=0}^{\infty} \frac{\psi_n(x;\alpha)}{n!} t^n$$

to derive the orthonormal polynomial basis $\{L_n(x;\alpha)\}$

$$L_n(x;\alpha) = \frac{1}{n!\alpha^{\overline{n}}} \psi_n(x;\alpha)$$

where $\alpha^{\overline{n}} = \alpha(\alpha+1)\ldots(\alpha+n-1)$.

Orthonormal Polynomial Bases

Example (Gamma Density)

In this case

$$L_n(x;\alpha) = \Gamma(n+\alpha)\sum_{m=0}^{n}\binom{n}{m}\frac{(-1)^m x^m}{\Gamma(m+\alpha)}$$

$L_n(x;\alpha)$ is the (generalized) *Laguerre polynomial of order n*.

If $\alpha = 1$, we have the recursion for $n \geq 1$

$$L_{n+1}(x) = \frac{1}{n+1}((2n+1-x)L_n(x) - nL_{n-1}(x))$$

with $L_0(x) = 1, L_1(x) = 1 - x$.

Orthonormal Polynomial Bases

Other orthogonal polynomial bases include

- Chebychev polynomials

$$p(x) = \frac{1}{\pi\sqrt{1-x^2}} \qquad -1 < x < 1$$

- Gegenbauer polynomials

$$p(x) = \frac{\Gamma(\alpha+1)}{\sqrt{\pi}\Gamma(\alpha+1/2)}(1-x^2)^{\alpha-1/2} \qquad -1 < x < 1$$

- Legendre polynomials

$$p(x) = \frac{1}{2} \qquad -1 < x < 1$$

Implication for Quadrature

Recall the quadrature approximation:

$$\int_{-\infty}^{\infty} g(x)p(x)\,dx \doteq \sum_{j=0}^{k} w_j g(x_j)$$

Provided the support of p is not finite, and provided that p has an associated orthonormal polynomial basis $\{\psi_n\}$, then this approximation can be made **exact** for integrands g that are themselves polynomials of order $2k+1$ or lower by choosing the points $x_0, \ldots, x_k$ to solve

$$\psi_{k+1}(x) = 0$$

with weights w_j determined by the $\{\psi_n\}$.

Frequentist example

Example (Variance of Normal median)

If $X_1, \ldots, X_n \sim \mathcal{N}(0,1)$ are a random sample, the sample median $\widetilde{X}$ has (for n odd) the pdf

$$n\binom{n-1}{m}\Phi(x)^m(1-\Phi(x))^m\phi(x)$$

where $m = \lfloor n/2 \rfloor$, and therefore the variance is

$$n\binom{n-1}{m}\int_{-\infty}^{\infty} x^2\Phi(x)^m(1-\Phi(x))^m\phi(x)\,dx = \int_{-\infty}^{\infty} g(x,n)\phi(x)\,dx$$

say. Using the central limit theorem and delta method

$$\sqrt{n}\widetilde{X} \xrightarrow{d} \mathcal{N}(0, \pi/2)$$

so for large n, the variance is approximately $1.5708/n$.

Frequentist example

Example (Variance of Normal median)

Using the Hermite polynomial basis formulation, the definition of the roots $x_0, x_1, \ldots, x_k$ and the fact that we have

$$\int_{-\infty}^{\infty} H_j(x)\phi(x)\,dx = 0$$

for each j, the weights $\boldsymbol{w} = (w_0, \ldots, w_k)^\intercal$ in the quadrature formula can be found by solving

$$\Psi(\boldsymbol{x})\boldsymbol{w} = \begin{pmatrix} \int_{-\infty}^{\infty} H_0(x)\phi(x)\,dx \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where the (i,j)th element of $\Psi(\boldsymbol{x})$ is $H_i(x_j)$, for $i, j = 0, \ldots, k$.

R implementation

There are functions in R that evaluate the orthonormal polynomials. For (univariate) Gauss-Hermite integration the function `gqz` in the `npmlreg` library.

Example (Variance of Normal median)

For different values of n and k, the following exact values of the variance of the median estimator were found.

	k					
n	10	20	50	100	1000	Exact
11	0.21018	0.14756	0.13717	0.13716	0.13716	0.13716
21	0.13492	0.10954	0.07455	0.07326	0.07326	0.07326
51	0.02228	0.05311	0.04496	0.03204	0.03054	0.03054
101	0.00075	0.01101	0.02858	0.02268	0.01549	0.01549

R Code for Gauss-Hermite integration

```
1  library(npmlreg)
2
3  g.func<-function(xv,nv){
4    mv<-floor(nv/2)
5    gv<-nv*choose(nv-1,mv)*xv^2
6           *(pnorm(xv)^mv)*((1-pnorm(xv))^mv)
7    return(gv)
8  }
9
10 n<-11
11 nk<-20
12 gh<-gqz(nk,minweight=1e-14)
13 x<-gh$location
14 w<-gh$weight
15 g<-g.func(x,n)
16 print(c(n,ik,sum(w*g)))
```

Higher dimensions

Gauss-Hermite integration extends to multivariate integration problems. The simplest extension utilizes the product approximation

$$\int_{-\infty}^{\infty} g(x_1, x_2) p(x_1, x_2)\, dx_1 dx_2 \doteq \sum_{i=0}^{k} \sum_{j=0}^{k} w_i w_j g(x_{1i}, x_{2j})$$

on a rectangular grid, where the weights are determined identically to the univariate case.

However, for general functions, the approximation breaks down in high dimensions

- in p dimensions, need p^{k+1} function evaluations.
- "high" translates as tens.