Mathematical Definitions

Ofir Shukron

1 P-values

Given a rejection region Γ in the form of $[c, \infty)]$, for the null hypothesis H_0, the p-value of an observed statistics $T = t$ is defined as

$$p - value(t) = min_{\{\Gamma; t \in \Gamma\}} \{Pr(T \in \Gamma | H_0 \quad true)\}$$

See definition in [4].

2 Type I and Type II errors

A type I and type II errors might occur in statistical hypothesis testing when hypothesizing about the observed null hypothesis. Type I error is accepting the alternative (or rejecting the null) when the null is correct, while type II is accepting the null when the alternative is correct.

3 Quadratic Forms

From Mathai [3](Chapters 1 and 4). For a $p \times 1$ random vector X with mean value $E(X) = \mu$ and $Cov(X) = E[(X - E(X))(X - E(X))'] = \Sigma > 0$, we set

$Y = \Sigma^{-0.5} X \Rightarrow E(Y) = \Sigma^{-0.5}\mu$ and $Cov(Y) = \Sigma^{-0.5} Cov(X) \Sigma^{-0.5} = I$ $Z = (Y - \Sigma^{-0.5}\mu) \Rightarrow E(Z) = 0$ and $Cov(Z) = I$ We can express the quadratic form $Q(X) = X'AX$ with the centralize variable Z and a symmetric positive definite matrix A as

$$Q(X) = X'AX = Y'\Sigma^{0.5} A \Sigma^{0.5} Y = (Z + \Sigma^{-0.5}\mu)'\Sigma^{0.5} A \Sigma^{0.5}(Z + \Sigma^{-0.5}\mu) \quad (1)$$

As an example, if we set $\mu = 0$, $A = I$, $Cov(X) = \Sigma = \sigma I$ and get $Q(X) = X'X$, $Y = \frac{X}{\sqrt{\sigma}}$, $Z = (\frac{X}{\sqrt{\sigma}} - 0)$, then

$$Q(X) = X'X = (\frac{X}{\sqrt{\sigma}})'(\sigma I)(\frac{X}{\sqrt{\sigma}}) = XX' \tag{2}$$

which is the normal form.

If P is a $p \times p$ orthogonal matrix which diagonalize $\Sigma^{0.5}A\Sigma^{0.5}$, that is

$$P'\Sigma^{0.5}A\Sigma^{0.5}P = diag(\lambda_1, \lambda_2, ..., \lambda_p)$$

with $P'P = I$, and λ_i are the eigenvalues of $\Sigma^{0.5}A\Sigma^{0.5}$. Then if $U = P'Z$, we have

$$Z = PU, \qquad E(U) = 0 \qquad Cov(U) = I$$

Then we can express the quadratic form of X by

$$Q(X) = (U + b)'diag(\lambda_1, \lambda_2, ..., \lambda_p)(U + b) \tag{3}$$

with $b = (P'\Sigma^{-0.5}\mu)'$

4 Spline interpolation

This section is mainly taken from [2]. The Newton from of the interpolation polynomial is defined using the divided differences as:

$$p_n(x) = \sum_{i=1}^{n}(x - \tau_1) \cdots (x - \tau_n)[\tau_1, \tau_2, ..\tau_n]g$$

where τ_i are the nodes of the function g to be interpolated, in which the polynomial $p_n(x)$ of order n agrees with it i.e $p_n(\tau_i) = g(\tau_i) \quad \forall i = 1..n$

5 the Heat equation $u_t = ku_xx$

Taken from [1]. The heat equation is a second order linear parabolic PDE describing the change of temperature in a surface with time. The constant k is called the conductivity . A fundamental source solution to the one-dimensional homogeneous equation is given by

$$u(x,t) = (4\pi kt)^{-0.5}\exp(-x^2/(4kt)) \quad t > 0, -\infty < x < \infty$$

the area under the graph of the solution is always 1. In general, to solve the equation we need to specify an initial condition (in terms of time) and two boundary conditions (in terms of space).

To begin to solve the heat equation, we search for product solutions of the form $u(x,t) = X(x)T(t)$, thus we perform the differentiation to obtain a function of time to equal a function of space, which is only possible if the two equal a constant c. Thus we arrive at two equations

$$T'(t) - kcT(t) = 0$$

$$X''(x) - cX(x) = 0$$

which their solution depends on the value of c

$$u(x,t) = e^{-\lambda^2 kt}(c_1 \sin(\lambda x) + c_2 \cos(\lambda x), \quad c = -\lambda^2 < 0$$

$$u(x,t) = e^{\lambda^2 kt}(c_1 e^{\lambda x} + c_2 e^{-\lambda x}), \quad c = \lambda^2 > 0$$

$$u(x,t) = c_1 x + c_2, \quad c = 0$$

If $u(x,t)$ is a solution of $u_t = ku_{xx}$ then

1. $v(x,t) = u(ax - x_0, a^2 t - t_0)$ is a solution of $v_t = kv_{xx}$ for any a, x_0, t_0
2. $v(x,t) = u(x, (k'/k)t)$ is a solution of $v_t = k'v_{xx}$

The solutions of the heat equations are **unique**, in the sense that if u_1, u_2 are solutions of $u_t = ku_{xx}$ with $0 \le x \le L, t \ge 0$, $u(0,t) = a(t), u(L,t) = b(t), u(x,0) = f(x)$, then $u_1 = u_2$

The stability, or change of solution in respect to small variation in the initial conditions, is seen through the **maximum principle**. If $u(x,t)$ is a C^2 solution of $u_t = ku_{xx}$ with $0 \le x \le L, t \ge 0$, $u(0,t) = a(t), u(L,t) = b(t), u(x,0) = f(x)$, with a, b, f a given C^2 functions with A, B, M the maximum temperature of $a(t), b(t), f(x)$ respectively and $\bar{M} = max(A, B, M)$. Then $u(x,t) \le \bar{M}$ in the whole range. The proof of the maximal principle can be found in [1] page 148.

For an **inhomogeneous heat equation**, with a heat source term $Q(x,t)$ we have

$$u_t = ku_{xx} + Q(x,t)$$

with B.C $u(-1,t) = u(1,t)$ and I.C $u(x,0) = f(x) = const$, we perform the following steps:

1. we first solve the homogeneous equation $u_t = ku_{xx}$ by separation of variables, to find the eigenfunctions $\sin(n\pi x)$ and eigenvalues $(n\pi)^2$, $n = 1..\infty$

2. then the solution can be expanded in terms of the eigenfunctions as

$$u = \sum_{n=1}^{\infty} c_n(t)\sin(n\pi x)$$

3. assume that the source term can be expanded as a Fourier series

$$Q(x,t) = \sum_{n=1}^{\infty} d_n \sin(n\pi x)$$

From the Fourier series theory we know that

$$d_n = \int Q(x,t)\sin(n\pi x)dx$$

but d_n must be a function of time and only a constant as a function of x, therefore setting $a_n(t) = d_n$ and assuming we can perform the integration

$$Q(x,t) = \sum_{n=1}^{\infty} a_n(t)\sin(n\pi x)$$

4. we now substitute the series expansion of u into the inhomogeneous equation to obtain

$$\sum_{n=1}^{\infty} c_n'(t)\sin(n\pi x) = -k\sum_{n=1}^{\infty}(n\pi)^2 c_n(t)\sin(n\pi x) + \sum_{n=1}^{\infty} a_n(t)\sin(n\pi x)$$

5. the Fourier coefficients must be equal, then we have the ODE

$$c_n'(t) = -k(n\pi)^2 c_n(t) + a_n(t) = -k\lambda_n c_n(t) + a_n(t)$$

6. we can solve the ODE by multiplying by the integration factor $e^{\lambda_n kt}$

The **Duhanel's principle** for inhomogeneous heat equations. suppose that $h(x,t) \in C^2 \quad 0 \leq x \leq L, t \geq 0$ then the inhomogeneous source problem $u_t - ku_{xx} = h(x,t)$ with $u(0,t) = u(L,t) = 0 \quad u(x,0) = 0$, is solved by first solving the related problem $v_t = kv_{xx}$ with $v(0,t;s) = v(L,t;s) = 0 \quad v(x,s;s) = h(x,s) \quad 0 \leq x \leq L, t \geq s$ in which the source term is translated into initial condition, and then performing the integration

$$u(x,t) = \int_0^t v(x,t-s;s)ds$$

For **inhomogeneous boundary conditions** we have to present the solution as a combination of a particular and a homogeneous solutions. So first we look at homogeneous problems in an infinite domain $u_t = ku_{xx} \quad -\infty \leq x \leq \infty \quad t \geq 0$ with the initial and boundary conditions: $u(x,0) = f(x)$.

We use the heat kernel, which is also the fundamental solution for the heat equation, namely

$$G(x,t) = \frac{1}{\sqrt{4\pi kt}} \exp(-\frac{x^2}{4kt})$$

The fundamental solution solve the initial value problem with singular initial data $G(x,t \to 0) = \delta(x)$ and it satisfies the heat equation. This function is exactly the Green's function. The Green's function $G(x,t;\zeta)$ is a solution with the initial singular data $\delta(x-\zeta)$

$$G(x,t;\zeta) = G(x-\zeta,t) = \frac{1}{\sqrt{4\pi kt}} \exp(-\frac{(x-\zeta)^2}{4kt})$$

The solution to the initial problem is then given by

$$u(x,t) = \int_{-\infty}^{\infty} G(x,t;\zeta)f(\zeta)d\zeta$$

References

[1] David Bleecker and George Csordas. *Basic partial differential equations.* CRC Press, 1992.

[2] Carl De Boor. A practical guide to splines. *Mathematics of Computation*, 1978.

[3] Arakaparampil M Mathai and Serge B Provost. *Quadratic forms in random variables: theory and applications*. M. Dekker New York, 1992.

[4] John D Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002.