Statistical Inference

Robert L. Wolpert
Institute of Statistics and Decision Sciences
Duke University, Durham, NC, USA

1. Random Samples & Generating Functions

Much of statistical theory is concerned with inference on the basis of a "sample" $X = (X_1, \cdots, X_n)$ of some number n of independent replicates, all from the same probability distribution with pdf or pmf $f(x \mid \theta)$; interest centers, in part, on what happens as n becomes large.

For any real-valued random variable X the complex-valued *characteristic function* (ch.f.)

$$\phi(\omega) \equiv \mathsf{E}[e^{i\omega X}]$$

is well-defined and satisfies $|\phi(\omega)| \le 1$, since $|e^{ir}|^2 = \cos^2 r + \sin^2 r = 1$ for all real numbers r. If X has finite mean μ and variance σ^2, the Lebesgue dominated convergence theorem justifies differentiation under the integral sign to compute

$$\begin{aligned}
\frac{d}{d\omega}\phi(\omega) &= \mathsf{E}[i\,X\,e^{i\omega X}] \\
\phi'(0) &= i\mathsf{E}[X] \\
\frac{d^2}{d\omega^2}\phi(\omega) &= \mathsf{E}[-\,X^2\,e^{i\omega X}] \\
\phi''(0) &= -\mathsf{E}[X^2]
\end{aligned}$$

so $\mu = -i\phi'(0)$ and $\sigma^2 = -\phi''(0) + \phi'(0)^2$ or, in terms of $\psi(\omega) \equiv \log \phi(\omega)$,

$$\mu = -i\psi'(0) \qquad \sigma^2 = -\psi''(0)$$

For an $\mathbb{R}^p$-valued random variable X the characteristic function is defined on $\mathbb{R}^p$ by

$$\phi(\omega) \equiv \mathsf{E}[e^{i\omega\cdot X}]$$

and satisfies the vector and matrix relations

$$\mu = -i\nabla\psi(0) \qquad \Sigma = -\nabla^2\psi(0)$$

Characteristic functions are just the Fourier transforms of distribution measures, so all the familiar results from Fourier analysis can be applied to them. For example, the Fourier inversion theorem lets us recover a distribution from its ch.f., we can relate smoothness of a ch.f. to the finiteness of moments of the distribution, etc.

Two other sorts of generating functions are commonly encountered: the **moment generating function** (MGF)

$$M(t) = \phi(-i\,t) = \mathsf{E}[e^{tX}]$$

derives its name from the property that

$$M^{(k)}(0) = \mathsf{E}[X^k]$$

when both sides exist, and in particular that $\mu = M'(0)$ and $\sigma^2 = M''(0) - M'(0)^2$ (or, better, that

$$\vec{\mu} = \nabla \log M(t)\big|_{t=0}, \qquad \Sigma = \nabla^2 \log M(t)\big|_{t=0};$$

we will use this property below, when studying Exponential Families), and the **generating function**

$$G(z) = \phi(-i\log z) = \mathsf{E}[z^X],$$

useful primarily for nonnegative integer-valued distributions since for these

$$\mathsf{P}[X = k] = G^{(k)}(1)/k!$$

Both these functions may fail to exist (or anyway to be finite) for thick-tailed distributions; the standard Cauchy distribution, for example, has MGF satisfying $M(t) = \infty,\ t \neq 0$.

2. Exponential Families

We will pay special attention to the **exponential family** in which each pdf takes the form

$$f(x \mid \theta) = \exp\left[\sum_{i=1}^{q} \eta_i(\theta) t_i(x) - B(\theta)\right] h(x)$$

(surprisingly many of the commonly-considered distributions can be written in this form, for suitable h, B, and $\{\eta_i, t_i\}_{i\le q}$), but will also consider other distributions. The likelihood function for a random sample of size n from the exponential family is

$$L(\theta) = \exp\left[\sum_{i=1}^{q} \eta_i(\theta) \sum_{j=1}^{n} t_i(x_j) - nB(\theta)\right]$$

It is often convenient to reparametrize exponential families to the *natural parameter* $\eta = \eta(\theta) \in \mathbb{R}^q$, leading (with $A(\eta(\theta)) \equiv B(\theta)$) to

$$f(x \mid \eta) = e^{\eta\cdot T(x) - A(\eta)}\, h(x)$$

Since any pdf integrates to unity we have

$$e^{A(\eta)} = \int_{\mathcal{X}} e^{\eta\cdot T(x)}\, h(x)\, dx$$

and hence can calculate the moment generating function (MGF) for the **natural statistic** $T(x) = \{t_1(x), \cdots, t_q(x)\}$ as

$$\begin{aligned}
M_T(s) &= \mathsf{E}\left[e^{s\cdot T(X)}\right] \\
&= \int_{\mathcal{X}} e^{s\cdot T(x)}\, e^{\eta\cdot T(x) - A(\eta)}\, h(x)\, dx \\
&= e^{-A(\eta)} \int_{\mathcal{X}} e^{(\eta+s)\cdot T(x)}\, h(x)\, dx \\
&= e^{A(\eta+s) - A(\eta)},
\end{aligned}$$

so we can find moments for the natural statistic by

$$\begin{aligned}
\mathsf{E}[T] &= \nabla \log M_T(0) &= \nabla A(\eta) \\
\mathsf{V}[T] &= \nabla^2 \log M_T(0) &= \nabla^2 A(\eta).
\end{aligned}$$

provided that η is an interior point of the *natural parameter space*

$$\mathcal{E} \equiv \{\eta \in \mathbb{R}^q : 0 < \int_{\mathcal{X}} e^{\eta\cdot T(x)}\, h(x)\, dx < \infty\}$$

and that $A(\cdot)$ is twice-differentiable near η.

3. Exponential Family Examples

$\mathsf{Be}(\alpha,\beta)$

$$
\begin{aligned}
f(x) &= \tfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad x \in (0,1) & T &= (\log x, \log 1{-}x)\\
B(\alpha,\beta) &= \gamma(\alpha)+\gamma(\beta)-\gamma(\alpha+\beta) & \eta &= (\alpha,\beta)\\
A(\eta) &= \gamma(\eta_1)+\gamma(\eta_2)-\gamma(\eta_1+\eta_2) & &\\
\nabla A(\eta) &= \begin{bmatrix}\psi(\eta_1)-\psi(\eta_1+\eta_2)\\ \psi(\eta_2)-\psi(\eta_1+\eta_2)\end{bmatrix} & \mathsf{E}T &= \begin{bmatrix}\psi(\alpha)-\psi(\alpha+\beta)\\ \psi(\beta)-\psi(\alpha+\beta)\end{bmatrix}\\
\nabla^2 A(\eta) &= \begin{pmatrix}\psi'(\eta_1)-c & -c\\ -c & \psi'(\eta_2)-c\end{pmatrix} & c &= \psi'(\eta_1+\eta_2)
\end{aligned}
$$

$\mathsf{Bi}(n,p)$

$$
\begin{aligned}
f(x) &= \tbinom{n}{x} p^x q^{(n-x)}, \quad x = 0 \ldots n & T &= x\\
B(p) &= -n \log q & \eta &= \log(p/q)\\
A(\eta) &= n \log(1+e^\eta) & p &= e^\eta/(1+e^\eta)\\
\nabla A(\eta) &= \tfrac{ne^\eta}{1+e^\eta} & \mathsf{E}T &= n\,p\\
\nabla^2 A(\eta) &= \tfrac{n\,e^\eta}{(1+e^\eta)^2} & I(p) &= n/pq
\end{aligned}
$$

$\mathsf{Ex}(\lambda)$

$$
\begin{aligned}
f(x) &= \lambda e^{-\lambda x}, \quad x > 0 & T &= x\\
B(\lambda) &= -\log\lambda & \eta &= -\lambda\\
A(\eta) &= -\log(-\eta) & &\\
\nabla A(\eta) &= -1/\eta & \mathsf{E}T &= 1/\lambda\\
\nabla^2 A(\eta) &= \eta^{-2} & I(\lambda) &= 1/\lambda^2
\end{aligned}
$$

$\mathsf{Ga}(\alpha,\lambda)$

$$
\begin{aligned}
f(x) &= \tfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1}\, e^{-\lambda x}, \quad x > 0 & T &= (\log x, x)\\
B(\alpha,\beta) &= \gamma(\alpha)-\alpha\log\lambda & \eta &= (\alpha,-\lambda)\\
A(\eta) &= \gamma(\eta_1)-\eta_1\log(-\eta_2) & &\\
\nabla A(\eta) &= \begin{bmatrix}\psi(\eta_1)-\log(-\eta_2)\\ -\eta_1/\eta_2\end{bmatrix} & \mathsf{E}T &= \begin{bmatrix}\psi(\alpha)-\log\lambda\\ \alpha/\lambda\end{bmatrix}\\
\nabla^2 A(\eta) &= \begin{pmatrix}\psi'(\eta_1) & -1/\eta_2\\ -1/\eta_2 & \eta_1/{\eta_2}^2\end{pmatrix} & I(\alpha,\lambda) &= \begin{pmatrix}\psi'(\alpha) & -1/\lambda\\ -1/\lambda & \alpha/\lambda^2\end{pmatrix}
\end{aligned}
$$

$\mathsf{Ge}(p)$

$$
\begin{aligned}
f(x) &= p\,q^x, \quad x = 0,1,2,\ldots & T &= x\\
B(p) &= -\log p & \eta &= \log q\\
A(\eta) &= -\log(1-e^\eta) & p &= 1-e^\eta\\
\nabla A(\eta) &= \tfrac{e^\eta}{1-e^\eta} & \mathsf{E}T &= q/p\\
\nabla^2 A(\eta) &= \tfrac{e^\eta}{(1-e^\eta)^2} & I(p) &= 1/p^2 q
\end{aligned}
$$

Exponential Family Examples (cont'd)

$\mathsf{IG}(a,b)$

$$
\begin{aligned}
f(x) &= a e^{-(a-bx)^2/2x}/\sqrt{2\pi x^3}, \quad x>0 & T &= (1/x, x)\\
B(a,b) &= -ab - \log a & \eta &= (-a^2/2, -b^2/2)\\
A(\eta) &= -2\sqrt{\eta_1\,\eta_2} - \tfrac{1}{2}\log(-2\eta_1) & a &= \sqrt{-2\eta_1}, \quad b = \sqrt{-2\eta_2}\\
\nabla A(\eta) &= \begin{bmatrix} \sqrt{\eta_2/\eta_1} - 1/2\eta_1 \\ \sqrt{\eta_1/\eta_2} \end{bmatrix} & \mathsf{E}T &= \begin{bmatrix} b/a + 1/a^2 \\ a/b \end{bmatrix}\\
\nabla^2 A(\eta) &= \tfrac{1}{2}\begin{pmatrix} \sqrt{\frac{\eta_2}{\eta_1{}^3}} + \frac{1}{\eta_1{}^2} & \frac{-1}{\sqrt{\eta_1\eta_2}} \\ \frac{-1}{\sqrt{\eta_1\eta_2}} & \sqrt{\frac{\eta_1}{\eta_2{}^3}} \end{pmatrix} & I(a,b) &= \begin{pmatrix} b/a + 2/a^2 & -1 \\ -1 & a/b \end{pmatrix}
\end{aligned}
$$

$\mathsf{NB}(\alpha,p)$

$$
\begin{aligned}
f(x) &= \tbinom{-\alpha}{x} p^\alpha\,(-q)^x, \quad x = 0,1,2,\ldots & T &= x\\
B(p) &= -\alpha\log p & \eta &= \log q\\
A(\eta) &= -\alpha\log(1-e^\eta) & p &= 1-e^\eta\\
\nabla A(\eta) &= \tfrac{\alpha e^\eta}{1-e^\eta} & \mathsf{E}T &= \alpha q/p\\
\nabla^2 A(\eta) &= \tfrac{\alpha e^\eta}{(1-e^\eta)^2} & I(p) &= \alpha/p^2 q
\end{aligned}
$$

$\mathsf{No}(\mu,\sigma^2)$

$$
\begin{aligned}
f(x) &= e^{-(x-\mu)^2/2\sigma^2}/\sqrt{2\pi\sigma^2} & T &= (x, x^2)\\
B(\mu,\sigma^2) &= \mu^2/2\sigma^2 + \tfrac{1}{2}\log\sigma^2 & \eta &= (\mu\sigma^{-2}, -\sigma^{-2}/2)\\
A(\eta) &= -\eta_1{}^2/4\eta_2 - \tfrac{1}{2}\log(-2\eta_2) & &\\
\nabla A(\eta) &= \begin{bmatrix} -\eta_1/2\eta_2 \\ \eta_1{}^2/4\eta_2{}^2 - 1/2\eta_2 \end{bmatrix} & \mathsf{E}T &= \begin{bmatrix} \mu \\ \mu^2+\sigma^2 \end{bmatrix}\\
\nabla^2 A(\eta) &= \begin{pmatrix} -1/2\eta_2 & \eta_1/2\eta_2{}^2 \\ \eta_1/2\eta_2{}^2 & -\eta_1{}^2/2\eta_2{}^3 + 1/2\eta_2{}^2 \end{pmatrix} & I(a,b) &= \begin{pmatrix} \sigma^{-2} & 0 \\ 0 & \sigma^{-4}/2 \end{pmatrix}
\end{aligned}
$$

$\mathsf{Po}(\lambda)$

$$
\begin{aligned}
f(x) &= \lambda^x e^{-\lambda}/x!, \quad x = 0,1,2,\ldots & T &= x\\
B(\lambda) &= \lambda & \eta &= \log\lambda\\
A(\eta) &= e^\eta & \lambda &= e^\eta\\
\nabla A(\eta) &= e^\eta & \mathsf{E}T &= \lambda\\
\nabla^2 A(\eta) &= e^\eta & I(\lambda) &= 1/\lambda
\end{aligned}
$$

$\mathsf{Pa}(\alpha,\beta)$

$$
\begin{aligned}
f(x) &= \beta\,\alpha^\beta/x^{\beta+1}, \quad x>\alpha & T &= \log x\\
B(\beta) &= -\log\beta - \beta\log\alpha & \eta &= -\beta\\
A(\eta) &= -\log(-\eta) + \eta\log\alpha & &\\
\nabla A(\eta) &= \log\alpha - 1/\eta & \mathsf{E}T &= \log\alpha + 1/\beta\\
\nabla^2 A(\eta) &= \eta^{-2} & I(\lambda) &= \beta^{-2}
\end{aligned}
$$