The fiducial method
What is it?

Peter McCullagh
University of Chicago

Chapel Hill, N.C.
Dec 5 2008

Fisher: the early years

- Fisher 1918: the correlation coefficient

Fisher: the early years

- Fisher 1918: the correlation coefficient
- Fisher 1924: Likelihood, sufficiency,...

Fisher: the early years

- Fisher 1918: the correlation coefficient
- Fisher 1924: Likelihood, sufficiency,...
- Fisher 1934: ancillarity...

Fisher: the early years

- Fisher 1918: the correlation coefficient
- Fisher 1924: Likelihood, sufficiency,...
- Fisher 1934: ancillarity...
- Fisher & K. Pearson

Fisher: the early years

- Fisher 1918: the correlation coefficient
- Fisher 1924: Likelihood, sufficiency,...
- Fisher 1934: ancillarity...
- Fisher & K. Pearson
- Fisher & J. Neyman

The classical Fisherian fiducial argument

If T is a statistic of continuous variation, and P is the probability that T should be less than any specified value, then $P = F(T, \theta)$. If we now give to P any particular value such as 0.95, we have a relationship between the statistic T and the parameter θ such that T is the 95% value corresponding to a given θ, and this relation implies the perfectly objective fact that in 5% of samples T will exceed the 95% value corresponding to the actual value in the population from which it is drawn. To any value of T there will moreover usually be a particular value of θ to which it bears this relationship; we may call this the "fiducial 5% value of θ" corresponding to T.

(Fisher SMRW 1930).

$F(T; \theta)$ is uniform on $(0, 1)$

The event $\{t : F(t; \theta) \geq 1 - \alpha\} = \{t : \theta \leq L_\alpha(t)\}$ has prob α

$L_\alpha(t)$ is the fiducial 5% point [for the parameter]

corresponding to the observation $T = t$

Fiducial distribution = confidence distribution

Zabell commenting on Fisher 1930

If the function $G(\alpha) = L_\alpha(t)$ is strictly decreasing in α, then its inverse $G^{-1}(\theta)$ is certainly a distribution function in the mathematical sense that $H(\theta) = 1 - G^{-1}(\theta)$ is continuous increasing with $H(-\infty) = 0$ and $H(\infty) = 1$; Fisher termed it the fiducial distribution of the parameter θ corresponding to the value $T = t$ and noted that it has the density $-\partial F(t, \theta)/\partial\theta$. Fisher regarded this result as supplying "definite information as to the probability of causes" and viewed the fiducial distribution as a probability distribution for θ in the ordinary sense.

(Zabell, Statistical Science 1992)

Conclusion:

To Fisher in 1930, fiducial density = confidence density

Fisher gave a general method for computing confidence limits for a real-valued parameter from a real-valued statistic. before the formal concept of a confidence interval existed.

Fréchet to Fisher: 8 Jan 1940

... and I would like to have your opinion on the following quotation of a paper by Deming and Birge. . . . *These particular values of σ and s are accordingly so related to each other that if σ were actually the s.d. of the parent population then there would be 19 chances in 20 that a sample drawn therefrom would have a s.d. as large or larger than s;*
<u>and conversely</u>, since s has actually been observed, there is only 1 chance in 20 that the s.d. of the parent population is as large or larger than σ.
. . . I have underlined <u>and conversely</u> because it seems to me that a logical confusion has there arisen. . . .
This equation holds good when σ being fixed, s is a random variable. . . .
Now in the second part of the quotation, the former value of this probability (computed before the trials) is considered also as valid when, s being observed, σ is considered as a random variable. As the whole proof of the formula is based on the first hypothesis and not the second one, I do not see why the formula should be still valid in these very different circumstances.

Fisher to Fréchet: 17 Jan 1940

I am very glad to have your letter of Jan 8 with the quotation from Birge and Deming. I find these writers often very obscure, but as they are obviously influenced by the form of argument for which I am responsible, I should like to make my own position, at least, clear to you. I enclose the paper on 'Inverse probability' ...

Fréchet to Fisher: 21 Jan 1940

I have read with great interest your paper 'Inverse probability'. I understand all the paper and agree with many of the statements, but I confess that I find here precisely the *same difficulty* as the one I pointed out in Deming and Birge's paper. The difficulty arises in a very short sentence which looks as obvious as the other ones but of which the meaning has a capital importance (as the words 'and conversely' in Deming and Birge):

Page 533: *we may express the relationship by saying that the true value of θ will be less than the fiducial 5% value corresponding to the observed value of T in exactly 5 trials in 100.* Now the 5% and the 5 trials in 100 refer to the same event, it is true, but the populations where these probabilities are computed are extremely different. ...

15 follow-up letters Jan – Apr 1940

Fréchet to Fisher: 14 June 1947
I send you separately a typewritten copy of my report on the 'Estimation of parameters' about the enquiry organized on that subject by the I.S.I. ... As I tried in pages 35–36, to explain your position, I think it best to send this copy to you, because if your answer would change my mind, I might still introduce corrections ....

Fisher to Fréchet: 21 June 1947
Thank you for your letter and courtesy in sending me your critical commentary. I shall not be inclined to argue the matter .... It seems to me obvious that serious mental obstacles must always have existed to the apprehension of a form of reasoning which seems to me new and valuable ... I should be glad if anyone reading my works should take what good they may find in them and make good use of it, and trouble themselves little about what they think to be false or defective.

Fréchet to Fisher: 18 Oct 1951
Vous savez que j'ai la plus grande admiration pour l'ensemble de votre oeuvres. vous savez aussi que je ne suis pas toujours d'accord avec vous sur les détails.

Parametric inference versus sample-space inferences

Context:

An index set $\mathbb{N} = \{1, 2, \ldots\}$

An infinite random sequence $Y = (Y_1, Y_2, \ldots)$: $Y \in \mathcal{S} = \mathcal{R}^\infty$

A probability distribution $(\mathcal{R}^\infty, \mathcal{F}, P)$: usually $\mathcal{F} = \mathcal{B}^\infty$ (Borel)

determined by finite-dimensional distns $(\mathcal{R}^n, \mathcal{F}_n, P_n)$

[Parametric] statistical model:

$\{(\mathcal{R}^\infty, \mathcal{F}, P_\theta) : \theta \in \Theta\}$ usually written $\{P_\theta\}$

...all defined on the same space $(\mathcal{R}^\infty, \mathcal{F})$

only one condition on Θ...

Sample: finite ordered subset $\{i_1, \ldots, i_n\}$ of $\mathbb{N}$

Observation: sample point $\mathbf{y} = (Y(i_1), \ldots, Y(i_n))$ in $\mathcal{R}^n$

Parametric inferences:

probabilistic statements concerning Θ (given observation)

Sample-space inferences:

probabilistic statements concerning $\mathcal{S}$ (given observation)

Parametric models: two 'typical' examples

Examples Ia, Ib:

Parameter space $\Theta = \mathcal{R} \times \mathcal{R}^+ = \{(\mu, \sigma) : \sigma > 0\}$

Example Ia: (iid Gaussian process) N_{μ,σ^2}

pick $\theta = (\mu, \sigma) \in \Theta$: (the true value)

generate $Y_1, Y_2, \ldots$ iid $N(\mu, \sigma^2)$

observe $\mathbf{y} = Y[n]$ on fixed sample $[n] = \{1, \ldots, n\}$

Prob distn: $P_{n,\theta}(A) = N_n(\mathbf{1}\mu, \sigma^2 I_n)(A)$

Example Ib: (Gosset process) $G_{\mu,\sigma}$

pick $\theta = (\mu, \sigma) \in \Theta$: (the true value)

generate $\epsilon_0 = \pm 1,\ \epsilon_1, \ldots$ independent $\epsilon_n \sim t_n$

$Y_1 = \mu,\ \ Y_2 = \mu \pm \sigma,\ \ Y_{n+1} = \bar{Y}_n + s_n \sqrt{(1 + 1/n)}\, \epsilon_{n-1}$

observe $\mathbf{y} = Y[n] = (Y_1, \ldots, Y_n)$ (not ϵ)

Parametric inference: goal concerns Θ

Sample-space inference: goal concerns sample-space events

such as $Y_{n+1} > 3.14$ or $\bar{Y}_\infty > \bar{Y}_n + 1.6$

Parametric models: Two atypical examples

Example IIa, IIb: Θ' is the space of prob distns ν on $\mathcal{R} \times \mathcal{R}^+$

Example IIa:

pick parameter $\nu \in \Theta'$: (the true value)
generate $\theta \sim \nu$ followed by $Y_1, \ldots,$ iid $N(\theta_1, \theta_2^2)$
observe $\mathbf{y} = Y[n]$ on fixed sample $[n] = \{1, \ldots, n\}$
Prob distns: $N_\nu(A) = \int N_{\mu,\sigma}(A)\, \nu(d\mu, d\sigma)$

$\theta \sim \nu$ not observed (not in the observation space)

Example IIb:

pick parameter $\nu \in \Theta'$: (the true value)
generate $\theta \sim \nu$ followed by $Y \sim G_\theta$ in $\mathcal{R}^\infty$
$Y_1 = \theta_1, \;\; Y_2 = \theta_1 \pm \theta_2, \;\; Y_{n+1} = \bar{Y}_n + s_n \sqrt{(1 + 1/n)}\epsilon_{n-1}$
Prob distn $G_\nu(A) = \int G_\theta(A)\, \nu(d\theta)$

Goal of parametric inference: to determine the parameter $\nu \in \Theta'$
Goal of sample-space inference: to predict $Y_{n+1}, \ldots$

The Gaussian pivotal model

Definition: the set of sequences $Y = (Y_1, Y_2, \ldots)$ such that
$\mu(Y) = \lim_{n\to\infty} \bar{Y}_n$ exists (finite) w.p.1
$\sigma^2(Y) = \lim_{n\to\infty} s_n^2 > 0$ exists (finite) w.p.1
$\epsilon_i = (Y_i - \mu(Y))/\sigma(Y)$ iid $N(0,1)$

Aim: Given $\mathbf{y} = Y[n]$ predict subsequent events
such as value of Y_{n+1} or $\sigma^2(Y)$ or $\bar{Y}_\infty \equiv \mu(Y)$

Is this a parametric model $\{(\mathcal{R}^\infty, \mathcal{F}, P_\theta) : \theta \in \Theta\}$?

Some properties of pivotal model:
closed under mixtures (i.e. convex set of prob distns)
closed under finite permutation $Y_{\pi(1)}, Y_{\pi(2)}, \ldots$
Includes Ia: iid $N_{\mu,\sigma}$, and all mixtures N_ν (spherically symmetric)
Includes Ib: $G_{\mu,\sigma}$, plus mixtures G_ν and permutations
... but these are only a few tips of the iceberg

Sample-space inference: general framework

A parametric model $\{P_\theta\}$ for a random sequence
(i) choose $\theta \in \Theta$ (or a distribution P_θ)
(ii) generate $Y \sim P_\theta$ in $\mathcal{R}^\infty$
(iii) observe $\mathbf{y} = Y[n]$ on a fixed sample $[n]$

Goal to predict subsequent values $Y_{n+1}, \ldots$
in the sense of conditional distribution given $Y[n]$

Unanimity condition:
$P_\theta(A \mid Y[n]) = P_{\theta'}(A \mid Y[n])$ for all θ, θ'

Unanimity automatic for all $n \geq 0$ if $\#\{P_\theta\} = 1$

Unanimity may hold for all $n \geq k > 0$ even if $\#\{P_\theta\} > 1$
Example: the Gosset process $G_{\mu,\sigma}$ with $n \geq 2$
Is this sufficient?

Procrustean strategies for sample-space inference

Procrustean goal: forced unanimity
to replace $\{P_\theta\}$ with a single 'representative' process Q
sample-space inferences: $Q(A \mid Y[n])$

Tibetan solution by acclamation: $Q = P_{\theta^*}$ is 'the chosen one'
declare θ^* chosen by 'higher authority' (GWB 2004)

Poor man's winner-take-all solution: $Q = P_{\hat{\theta}_n}$
Breaks the rules by allowing Q to depend on observation
Advantages: sometimes very simple, sometimes not so bad
Disadvantages: only works for $n \geq k$; incoherence;...

Bayes solution: Replace $\{P_\theta\}$ with a mixture $Q = P_\nu$
(proportional representation as determined by ν)
Advantages: works for all $n \geq 0$; good theoretical properties...
Disadvantages: computation; principles for choosing ν;...

Remark: For models IIa & IIb, mixture solution = Tibetan solution

Fiducial strategy for sample-space inference

Pivotal models:

(i) Group $\mathcal{G}$ acting on $\mathcal{R}^\infty$ and on Θ

$$P_{g\theta}(gA) = P_\theta(A) \text{ for } g \in \mathcal{G}$$

(ii) Action on $\mathcal{R}^\infty$ is component-wise

$$(y_1, y_2, \ldots) \mapsto (gy_1, gy_2, \ldots)$$

$\mathcal{F} \subset \mathcal{B}^\infty$ is the class of $\mathcal{G}$-invariant events

(iii) $A \in \mathcal{F}$ implies $P_\theta(A) = P_{\theta'}(A)$ all θ, θ'

[If the action of $\mathcal{G}$ on Θ is transitive, then (iii) follows from (i).]

Fiducial solution (focusing on points of agreement):

Replace $\{(\mathcal{R}^\infty, \mathcal{B}^\infty, P_\theta)\}$ with $Q = (\mathcal{R}^\infty, \mathcal{F}, P_\theta)$

Use conditional distn $Q(A \,|\, \mathcal{F}_n)$ for sample-space inferences

events $A \in \mathcal{F}$, $\qquad \mathcal{F}_n = \mathcal{F} \cap \mathcal{B}^n$

Choice of representative depends on $\mathcal{G}$ but not on observation

Example: Gaussian pivotal model

$\{P_\theta\}$: The set of sequences $Y = (Y_1, Y_2, \ldots)$ such that
$(Y_i - \mu(Y))/\sigma(Y)$ iid $N(0,1)$

$\mathcal{G}$ = location-scale group acting component-wise
$\mathcal{F} \subset \mathcal{B}^\infty$ is $\mathcal{G}$-invariant σ-field

Fiducial or $\mathcal{G}$-invariant solution:
Replace set $\{(\mathcal{R}^\infty, \mathcal{B}^\infty, P_\theta)\}$ with $Q = (\mathcal{R}^\infty, \mathcal{F}, P_\theta)$

$\mathcal{F}$ generated by the Borel random variables
$T_{n+1} = (Y_{n+1} - \bar{Y}_n)/s_n) \sim t_{n-1}(0, 1 + 1/n)$ independent for $n \geq 2$
$\implies Q((\bar{Y}_\infty - \bar{Y}_n)/s_n \in B \mid Y[n]) = t_{n-1}(0, 1/n)(B)$ for $n \geq 2$

Fiducial extension: replace Q with $\{G_\nu\}$ on Borel sets
$G_\nu(\bar{Y}_\infty \in B \mid Y[n]) = t_{n-1}(\bar{y}_n, s_n^2/n)(B)$ for $n \geq 2$ only

Max likelihood solution $= G_{\hat{\mu},\hat{\sigma}}$
Bayes solution = ?

Further examples of 'ordinary' parametric models

IID real-valued:

Θ is the set of distributions on $(\mathcal{R}, \mathcal{B})$

$Y_1, \ldots$ iid θ; $\quad P_{n,\theta} = \theta^n$ (product distribution)

Wishart tree model:

$\Theta = \{\Sigma : \Sigma_{ij} \geq \min(\Sigma_{ik}, \Sigma_{jk}) \geq 0\}$ (rooted trees)

Observation: a matrix $S \sim \mathcal{W}_n(\Sigma)$ such that $E(S) = \Sigma$

goal of parametric inference: to estimate $\Sigma \in \Theta$

Stationary AR1: parameter (σ, ρ)

$P_n = N_n(0, \Sigma)$; $\Sigma_{ij} = \tau^2 \rho^{|i-j|}$; $\quad |\rho| < 1$

$Y_1 \sim N(0, \sigma^2/(1-\rho^2))$; $Y_{t+1} = \rho Y_t + \sigma \epsilon_t$; $\quad \epsilon$ iid $N(0,1)$

AR1: parameter (α, σ, ρ)

$Y_1 = \alpha$; $Y_{t+1} = \rho Y_t + \sigma \epsilon_t$; ϵ iid $N(0,1)$

Goals...

Linear model with covariates X and block factor:

Block factor: $B(i,j) = 1$ if $i \sim j$ in same block; zero otherwise

Model distribution: $P_n = N(X\beta, \sigma_0^2 I_n + \sigma_1^2 B)$

$Y(i) = \beta' x(i) + \eta(i)$ where $\eta(i) = \epsilon(i) + \epsilon'(b(i))$; ϵ, ϵ' indep

Generalized Gaussian distribution $N_n(\mu, \Sigma, \mathcal{K})$ in $\mathcal{R}^n$

Parameters μ, Σ and subspace $\mathcal{K} \subset \mathcal{R}^n$ (kernel)
such that $N_n(\mu, \Sigma, 0) \equiv N_n(\mu, \Sigma)$.
$\mu \cong \mu + \mathcal{K}, \quad \xi' \Sigma \xi > 0$ for $\xi \in \mathcal{K}^0$

Q1: On what subsets of $\mathcal{R}^n$ is $N_n(\mu, \Sigma, \mathcal{K})$ defined?
A1: Only on invariant Borel sets A such that $A + \mathcal{K} = A$;
(Borel subsets of $\mathcal{R}^n/\mathcal{K}$)
$A + B = \{a + b : a \in A, b \in B\}$;
$x + \mathcal{K}$ is the orbit(coset) of x by $\mathcal{K}$

Q2: What is the value $N_n(\mu, \Sigma, \mathcal{K})(A)$ for such a set $A \subset \mathcal{R}^n/\mathcal{K}$?
A2: $N_n(\mu, \Sigma, \mathcal{K})(A) = N_n(\mu, \Sigma)(A)$

Q3: Duh! What is the point of that?
A3: Not immediately obvious, but... patience...

Generalized Gaussian distribution (contd)

If $L\colon \mathcal{R}^n \to \mathcal{R}^{n-k}$ has kernel $\ker(L) = \mathcal{K}$, then
$Y \sim N_n(\mu, \Sigma, \mathcal{K}) \Leftrightarrow LY \sim N(L\mu, L\Sigma L')$

Q4: What is the likelihood function in the model $N(\mu, \Sigma, \mathcal{K})$?
A4: $l(\mu, \Sigma; y) = -\frac{1}{2}(y-\mu)'WQ(y-\mu) + \frac{1}{2}\log \mathrm{Det}(WQ)$
$W = \Sigma^{-1}, \quad Q = I - K(K'WK)^{-1}K'W;$
K a basis for $\mathcal{K}$, $\quad WQ = L'(L\Sigma L')^{-1}L$

Q5: Connection with REML?
A5: Yes, if $\mathcal{K} = \mathrm{span}(X)$

Q6: Are such models widely used?
A6: Yes: REML for variance-component estimation;
conditional distribution for spline smoothing, spatial Kriging,...
fiducial models

Q6: What is the conditional distribution of Y_{n+1} given $\mathbf{y}^{(n)}$?
A6: The question needs to be re-phrased...

Generalized Gaussian process $Q = N(0, I, \mathbf{1})$

$\mathbf{1} = \mathrm{span}(1, 1, 1, \ldots)$
$N(0, 1, \mathbf{1})$ exchangeable process defined on invariant sets
Finite-dimensional distributions $Q_n = N_n(0, I_n, \mathbf{1})$ equivalent to
- $Y_i - Y_j = \epsilon_i - \epsilon_j$ where ϵ is iid Gaussian
- $Y_{n+1} - \bar{Y}_n \sim N(0, (1 + 1/n))$ indep for $n = 1, 2, \ldots$
- $\implies Q(Y_{n+1} - \bar{Y}_n \in A \,|\, \mathbf{y}^{(n)}) = N(0, 1 + 1/n)(A)$
- $\implies Q(\bar{Y}_\infty - \bar{Y}_n \in A \,|\, \mathbf{y}^{(n)}) = N(0, 1/n)(A)$

Two Borel extensions of Q:
- IID extension P_μ: $Y_i = \mu + \epsilon_i \sim N(\mu, 1)$
  - also mixtures $P_\nu = \int P_\mu \, \nu(d\mu)$ of iid Gaussian processes
- Gosset extension G_μ: $Y_i = \mu + \epsilon_i - \epsilon_1$
  - also mixtures G_ν of Gosset processes

In all cases $Y_i - \bar{Y}_\infty = \epsilon_i \sim N(0, 1)$ (iid)

Inferential statements about $\bar{Y}_\infty$

From the generalized Gaussian processes $Q_n = N(0, I_n, \mathbf{1})$
Given initial sequence $\mathbf{y}^{(n)} = (y_1, \ldots, y_n)$ with $n \geq 1$,
$Q(\bar{Y}_\infty - \bar{Y}_n \in A \,|\, \mathbf{y}^{(n)}) = N(0, 1/n)(A)$
(no useful inferential statements about $\bar{Y}_\infty$ if $n = 0$!)

IID extension $P_\mu(\bar{Y}_\infty \in A \,|\, \mathbf{y}^{(n)}) = \begin{cases} 1 & \mu \in A \\ 0 & \text{otherwise.} \end{cases}$

Exchangeable extension: $P_\nu(\bar{Y}_\infty \in A \,|\, \mathbf{y}^{(n)}) =$??

Gosset extension: $G_\mu(\bar{Y}_\infty \in A \,|\, \mathbf{y}^{(n)}) = N(\bar{y}_n, 1/n)(A)$
(independent of μ provided that $n \geq 1$).

Smoothing-spline models

Context: a regression model with real-valued x

Sample covariate configuration $\mathbf{x} = (x_1, \ldots, x_n)$

$K[\mathbf{x}]_{i,j} = \{-|x_i - x_j|\}, \quad \mathcal{K} = \mathrm{span}(\mathbf{1}, \mathbf{x})$ given

Model parameters σ_0^2, σ_1^2

Distributions $N(0, \sigma_0^2 I_n + \sigma_1^2 K[\mathbf{x}], \mathcal{K})$

Another pivotal model

IID model:

Parameter space Θ: set of cts distributions F on $(\mathcal{R}, \mathcal{B})$

Model: $Y_1, Y_2, \ldots$ iid F

$\hat{F} = \hat{F}_n$ not in Θ

Pivotal model:

set of inf exch sequences $Y_1, \ldots$ (distinct)

set of iid mixtures with $\nu(\Theta) = 1$

Fiducial solution for sample-space inference

$\mathcal{G}$ is group of monotone invertible maps $\mathcal{R} \to \mathcal{R}$

$\mathcal{F}$ is σ-field generated by ranks

$Q(Y_{n+1} \in A \,|\, \mathcal{F}_n) =$??