STSCI 4780/5780: Propagating uncertainty

Tom Loredo, CCAPS & SDS, Cornell University

Parametric modeling — Classes of problems

Single-model inference

Context = choice of single model (specific i)
Parameter estimation: What can we say about θ_i or $f(\theta_i)$?
Prediction: What can we say about future data D'?

Multi-model inference ("M-closed")

Context = $M_1 \vee M_2 \vee \cdots$
Model comparison/choice: What can we say about i?
Model averaging:

- *Systematic error*: $\theta_i = \{\phi, \eta_i\}$; ϕ is common to all
  What can we say about ϕ w/o committing to one model?
- *Prediction*: What can we say about future D', accounting for model uncertainty?

Model checking ("M-open")

Premise = $M_1 \vee$ "all" alternatives
Is M_1 adequate? (predictive tests, calibration, robustness)

Propagating uncertainty

Often the parameters that most directly or simply allow us to model the data are not the quantities we are ultimately interested in.

- To model the data, I need extra (uncertain) parameters beyond those of interest to me—a background level, a noise scale, a calibration factor. What do I know about the parameters of interest? $\rightarrow$ *Marginalization over nuisance parameters*

- I have *two or more* rival parametric models for the available data. How strongly does the evidence favor one model over competitors, *accounting for parameter uncertainty*? $\rightarrow$ *Model comparison*

- I model available data, D, using a parametric model. What can I say about future data, D'? $\rightarrow$ *Prediction*

- I model binary outcome data in terms of the success probability, α. What have I learned about the failure probability, $\beta \equiv 1 - \alpha$? Or about the odds favoring success, $\omega \equiv \frac{\alpha}{1-\alpha}$?
  $\rightarrow$ *Change of variables*

The LTP will play a key role in addressing these problems.

Change of variables: Binomial inference

Recall the binomial inference problem, using success count data, n, and a flat/uniform prior:

$$\pi(\alpha) = 1; \qquad \mathcal{L}(\alpha) = \frac{N!}{n!(N-n)!}\alpha^n(1-\alpha)^{N-n}$$

$$\rightarrow p(\alpha|n) = \frac{(N+1)!}{n!(N-n)!}\alpha^n(1-\alpha)^{N-n}$$

What does this tell us about $\beta \equiv P(\text{failure}) = 1-\alpha$?

It's tempting to just *rewrite it substituting $1-\beta$ for α*:

$$\pi(\beta) = 1; \qquad \mathcal{L}(\beta) = \frac{N!}{n!(N-n)!}(1-\beta)^n\beta^{N-n}$$

$$\rightarrow p(\beta|n) = \frac{(N+1)!}{n!(N-n)!}(1-\beta)^n\beta^{N-n}$$

This appears to have worked. It did, *but only by accident!*

What do the data tell us about the *odds*,

$$\omega \equiv \frac{\alpha}{1-\alpha}, \quad \text{with } \omega \in [0, \infty]$$

Try parameter substitution:

$$\omega - \omega\alpha = \alpha \quad \rightarrow \quad \omega = \alpha(1+\omega) \quad \rightarrow \quad \alpha = \frac{\omega}{1+\omega}$$

We're already in trouble with the flat (w.r.t. α) prior!

$$\pi(\omega) = 1 \quad \rightarrow \quad \int_0^\infty d\omega\ \pi(\omega) = \infty$$

The swap-in posterior also can be improper (not normalizable):

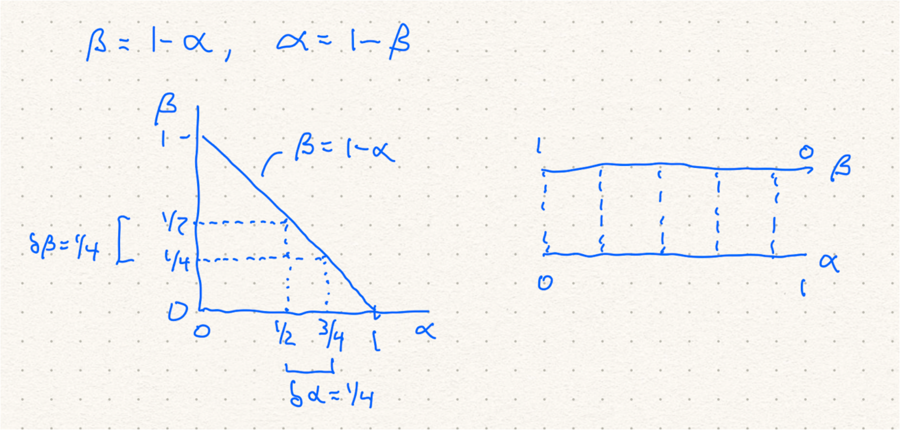
$$\alpha^n (1-\alpha)^{N-n} \quad \rightarrow \quad \left(\frac{\omega}{1+\omega}\right)^n \left(\frac{1}{1+\omega}\right)^{N-n}$$

For $N = 2$ and $n = 1$, we expect equal probability for $\omega < 1$ and $\omega > 1$, but the integral of $\omega/(1+\omega)^2$ diverges logarithmically for large ω

Why simple variable substitution (usually) fails

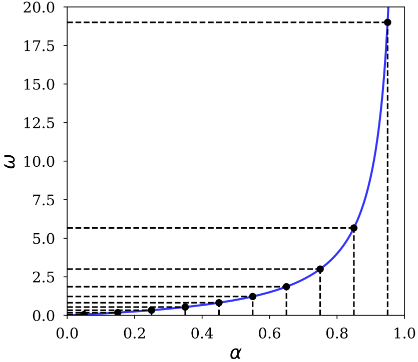
Key insight: A PDF for x is a probability *per unit x*

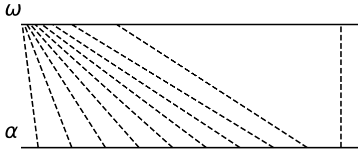
For $\alpha \leftrightarrow \beta$, equal α intervals corresond to equal β intervals.



Not so for $\alpha \leftrightarrow \omega$!

$$\omega \equiv \frac{\alpha}{1-\alpha} \qquad\qquad \alpha = \frac{\omega}{1+\omega}$$





A density (probability per unit interval) that is constant for α must *decrease* with increasing ω in order to assign the same probability to corresponding intervals.

Univariate change of variables

Recall the definition of a PDF for x:

$$P(x_* \in [x, x + \mathrm{d}x] \,|\, \ldots) = f(x)\,\mathrm{d}x \quad \text{for small } \mathrm{d}x$$

Let $y = Y(x)$, with a *one-to-one* function $Y(x)$, so y is a full, unique relabeling of the hypotheses labeled by x

There is a PDF for y:

$$P(y_* \in [y, y + \mathrm{d}y] \,|\, \ldots) = g(y)\,\mathrm{d}y \quad \text{for small } \mathrm{d}y$$

What $g(y)$ assigns probabilities to y intervals consistent with the probabilities $f(x)$ assigns to the corresponding x intervals?

We'll need the inverse map, from y to x: $x = X(y)$

Consistency condition: Require $f(x)$ and $g(y)$ to assign the same (small) probability to *corresponding* intervals δy and δx:

$$g(y)|\delta y| = f(x)|\delta x|$$

We want to relate δx and δy so that

$$[x, x+\delta x] \Longleftrightarrow [y, y+\delta y]$$

For the left boundary, set $x = X(y)$. For the right boundary:

$$\begin{aligned} x + \delta x &= X(y+\delta y) \\ X(y) + \delta x &\approx X(y) + X'(y)\delta y \\ \rightarrow \quad \delta x &= X'(y)\delta y \end{aligned}$$

The consistency cond'n becomes $g(y)|\delta y| = f[X(y)] \times |X'(y)\delta y|$, so

$$\boxed{g(y) = f[X(y)] \times |X'(y)|}$$

Mnemonic: $g(y)\,\mathrm{d}y = f(x)\,\mathrm{d}x \quad \rightarrow \quad g(y) = f(x)|\,\mathrm{d}x/\,\mathrm{d}y|$

Two examples

Take $x = \alpha, y = \beta$, so

$$Y(x): \quad \beta(\alpha) = 1 - \alpha$$
$$X(y): \quad \alpha(\beta) = 1 - \beta$$

What would a flat prior, $f(\alpha) = 1$, correspond to for β?

$$\begin{aligned} g(\beta) &= f(\alpha(\beta))\left|\frac{\mathrm{d}\alpha(\beta)}{\mathrm{d}\beta}\right| \\ &= f(1-\beta) \times 1 \\ &= 1 \end{aligned}$$

which *happens* to equal the substitution result, $f(1-\beta)$ (for *any* prior, $f(\alpha)$).

Take $y = \ln x$, so $x = e^y$. If we specify $f(x)$ as the PDF for x, what is the corresponding PDF for y, denoted $g(y)$?

$$\begin{aligned} Y(x) &= \ln x \\ X(y) &= e^y \end{aligned}$$

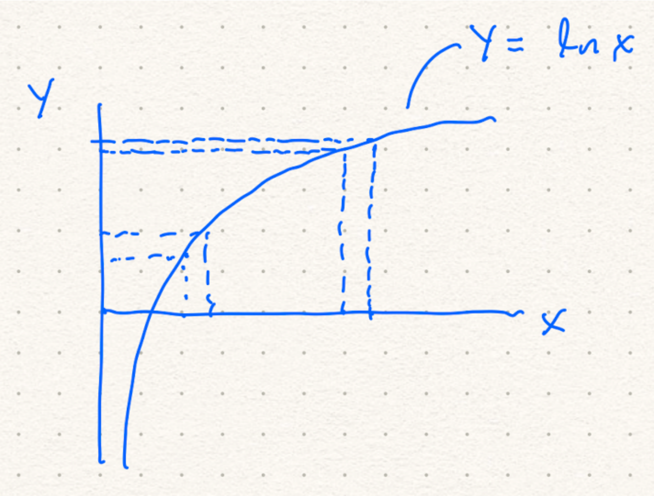
so

$$\begin{aligned} g(y) &= f(X(y))\left|\frac{\mathrm{d}X(y)}{\mathrm{d}y}\right| \\ &= f\left(e^y\right) \times e^y \end{aligned}$$

Suppose $f(x) = C/x$ (like some of our priors for a Poisson rate or Gaussian σ). Then:

$$g(y) = \frac{C}{e^y} \times e^y = 1$$

This is what is meant by saying the C/x prior is "log flat."



Y = ln x
Y
x

PDFs and expectations

PDFs are probability *densities*, not probabilities. We use them *inside integrals* to get probabilities and expectation values:

$$P(\theta \in [\theta_l, \theta_u] | \ldots) = \int_{\theta_l}^{\theta_u} \mathrm{d}\theta \; p(\theta|\ldots)$$

$$\mathbb{E}(x) = \int \mathrm{d}x \; x \, p(x|\ldots)$$

$$\mathbb{E}(g(x)) = \int \mathrm{d}x \; g(x) \, p(x|\ldots)$$

Note that

$$P(\theta \in [\theta_l, \theta_u] | \ldots) = \int \mathrm{d}\theta \; h(\theta) \, p(\theta|\ldots) = \mathbb{E}(h(\theta))$$

with $h(x) =$ a unit-height "box function" over $[\theta_l, \theta_u]$.

PDFs are mainly useful for computing *expectations*.

Expectations are functionals

In analysis (the branch of math), a *functional*, F, is a mapping from an *entire function*, $u(x)$, to a real number: $f = F[u]$ (note the square brackets).

Simple examples:

$$\text{Evaluation:} \quad F_c[u] = u(c)$$

$$\text{Area:} \quad F_{l,u}[u] = \int_l^u \mathrm{d}x \; u(x)$$

$$\text{Kernel:} \quad F_k[u] = \int \mathrm{d}x \; k(x) \; u(x)$$

These are all *linear functionals*: $F[a \times u] = aF[u]$ (for const. a).

Riesz representation theorem: Bounded linear functionals can always be represented as an integral over a kernel, like $F_k[\cdot]$, above.

Expectations are linear functionals, with PDFs playing the role of a kernal.

The Dirac δ function

What PDF gives us the evaluation functional? I.e., what PDF $p(x)$ communicates "the value of x is C"?

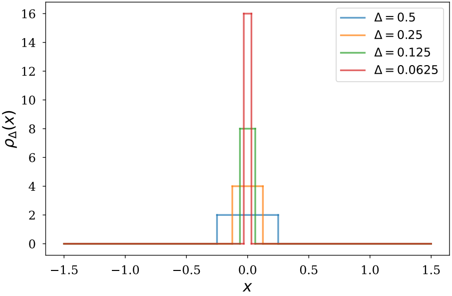
Recall from Lec05: The continuum is *tricky!* When θ_* may lie in a finite interval:

- $P(\theta_* = C|\ldots) = 0$ (this corresponds to $\mathrm{d}\theta = 0$)
- $P(\theta_* \in \mathbb{Z}|\ldots) = 0$, where $\mathbb{Z} =$ set of integers
- $\ldots$

This follows for any *finite, smooth* PDF, because multiplying $p(x)\,\mathrm{d}x$ vanishes in the limit $\mathrm{d}x \to 0$ as long as $p(x)$ is finite.

A PDF representing evaluation cannot be a finite, smooth function.

Build a "point mass at 0" PDF as a *limit* of finite PDFs that concentrate around a point:



$$\rho_\Delta(x) = \begin{cases} \frac{1}{\Delta} & \text{for } x \in [-\Delta/2, \Delta/2] \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\int_{-\infty}^{\infty} \mathrm{d}x \, \rho_\Delta(x) = \int_{-\Delta/2}^{\Delta/2} \mathrm{d}x \, \frac{1}{\Delta} = \frac{1}{\Delta} \times x \Big|_{-\Delta/2}^{\Delta/2} = 1$$

Blackboard work here!

Propagating uncertainty with δ functions

If we can compute $p(\theta|\ldots)$ for some scalar parameter θ, but we are interested in $f = F(\theta)$, we've established that

$$p(f|\ldots) = p(\Theta(f)|\ldots)\,|\Theta'(f)|,$$

where $\theta = \Theta(f)$ is the inverse map between f and θ

If θ is a vector, and f is a vector with the same number of components,

$$p(f|\ldots) = p(\theta|\ldots)\left|\frac{\partial\Theta(f)}{\partial f}\right|,$$

where the last factor is the Jacobian determinant of the transformation

Often we are interested in $F(\theta)$ that is lower-dimensional than θ, e.g., a scalar function of multiple parameters:

- The difference between two parameters, $s = r - b$
- The ratio of two parameters, $\rho = \theta/\psi$
- Some complex function: planet mass $M_p = f(K, \tau, e)$

The univariate case can be recast using the LTP and δ functions:

$$\begin{aligned} p(f) &= \int d\theta\, p(f,\theta) \\ &= \int d\theta\, p(\theta)\, p(f|\theta) \\ &= \int d\theta\, p(\theta)\, \delta[f - F(\theta)] \end{aligned}$$

Using the definition of δ as a limit of a narrow "hat" PDF,

$$\delta[f - F(\theta)] = \delta[\theta - \Theta(f)]\ |\Theta'(f)|$$

$$\begin{aligned} \rightarrow \quad p(f) &= \int d\theta\, p(\theta)\, \delta[\theta - \Theta(f)]\ |\Theta'(f)| \\ &= p[\theta = \Theta(f)]\ |\Theta'(f)| \end{aligned}$$

This reproduces our earlier result, but now in a generalizable setting

Two-parameter case: Parameters (θ, ψ) with $f = F(\theta, \psi)$

$$\begin{aligned} p(f) &= \int d\theta \int d\psi \, p(f, \theta, \psi) \\ &= \int d\theta \int d\psi \, p(\theta, \psi) \, p(f|\theta, \psi) \\ &= \int d\theta \int d\psi \, p(\theta, \psi) \, \delta[f - F(\theta, \psi)] \end{aligned}$$

Two ways this may be computed:

- Pick one of θ or ψ, find the inverse map, and use $\delta[\theta - \Theta(f, \psi)]$ or $\delta[\psi - \Psi(f, \theta)]$
- Monte Carlo: Draw samples $\{(\theta_i, \psi_i)\}$ from $p(\theta, \psi)$; compute $\{f_i\}$ with $f_i = F(\theta_i, \psi_i)$