Isaac's probability function, v.6

Pere d'Isaac

30 June, 2023 (at *11:54*)

Sampling a probability-density

Given a prob-density $D()$, how do we produce a D-random number?

The corresponding CDF (Cumulative Distribution Fnc) is

$$F(x) := \int_{-\infty}^{x} D \overset{\text{specifically}}{=\!=} \int_{(-\infty,x]} D.$$

So for each $\mathbf{x_0}$ real:

$*$: *A D-random point $q \in \mathbb{R}$ has the property that $\mathsf{P}_D(q \le \mathbf{x_0})$ equals $F(\mathbf{x_0})$.*

Suppose $D{>}0$ on $\mathbb{R}$. Then F is continuous and strictly-increasing, mapping $\mathbb{R} \hookrightarrow\!\!\!\rightarrow (0,1)$. Function-inverse F^{-1} maps $(0,1)$ onto $\mathbb{R}$.

Fix $\mathbf{y_0} \in (0,1)$. Then $\mathbf{y_0}$ is the probability –*using the uniform-distribution on* $(0,1)$– that a chosen point y in $(0,1)$ satisfies $\boxed{y \le \mathbf{y_0}}$.

With $q := F^{-1}(y)$ and $\mathbf{x_0} := F^{-1}(\mathbf{y_0})$, note

$$y \le \mathbf{y_0} \iff q \le \mathbf{x_0},$$

since F^{-1} is strictly increasing. Hence the probability that $q \le \mathbf{x_0}$ equals $\mathbf{y_0} \overset{\text{note}}{=\!=} F(\mathbf{x_0})$, as desired in $(*)$.

1a: ALGORITHM A:. Use the uniform-distribution to pick $y \in (0,1)$, then **Return $F^{-1}(y)$**. □

Possible alternative. Suppose density $D()$ is zero outside of some interval $[\mathsf{L}, \mathsf{H}]$.

If F^{-1} is easy to compute, then use (1a). However, if F^{-1} is difficult [or if the density is zero on intervals inside $[\mathsf{L}, \mathsf{H}]$, so there *is* no global F^{-1}], one can generate a random number directly from D, with the following *not-terribly-efficient* algorithm.

Let Max here mean the maximum [technically, ***supremum***] of $D(x)$ over all $x \in [\mathsf{L}, \mathsf{H}]$.

Fix any number $M \ge \text{Max}$.

1b: ALGORITHM B:. Pick x uniformly in $[\mathsf{L}, \mathsf{H}]$ and y uniformly in $[0, M]$. *If* $y \le D(x)$, then **Return x**; *else* go back and pick new x, y. □

For this algorithm, our D *need not* be normalized. E.g, replacing $D()$ by $\frac{1}{3}D()$ would not change the induced prob-measure. [But it would increase the expected number of iterations before an x is returned.]

The expected # of iterations can be reduced, by lowering M to equal Max.

Technical. A collection $\mathfrak{C}$ of $\mathbb{R}$-subsets is a ***σ-algebra*** if $\mathfrak{C}$ is sealed under *complement* and *countable union*. I.e,

$$B \in \mathfrak{C} \implies [\mathbb{R} \smallsetminus B] \in \mathfrak{C} \quad \text{and}$$

$$B_j \in \mathfrak{C} \implies \Big[\bigcup_{j=1}^{\infty} B_j\Big] \in \mathfrak{C}.$$

The ***Borel σ-algebra*** is the smallest σ-algebra owning all the open intervals.

A ***probability measure*** on $\mathbb{R}$ assigns a probability $\mathsf{P}(B)$ to each Borel set B. Its CDF (Cumulative Distribution Fnc) is $F(x) := \mathsf{P}\big((-\infty, x]\big)$.

A probability measure is ***atomic*** if $\exists x \in \mathbb{R}$ with $\mathsf{P}(\{x\})$ positive. A probability has a ***density*** (PDF) IFF it is non-atomic. Moreover, its density is the derivative of its CDF. [THEOREM: A CDF is differentiable "almost everywhere", that is, off of a probability-zero set.] □

The Problem. Fix <u>positive</u> reals $\mathsf{L} < \mathsf{H}$ (Low<High) and biases α, β. We seek a single-humped prob-measure on $\mathbb{R}$ which we condition on interval $[\mathsf{L}, \mathsf{H}]$, then emit a random #. There is a *seller* who will (grumpily) accept as little as L for his MacGuffin, and a *buyer* who will pay as much as H; both values secret.

The biases α, β are the fair-market-opinions of the *seller* and *buyer*, respectively, of MacGuffin's value.

[Originally, I misunderstood the biases, thinking that Isaac wanted **affine**-invariance w.r.t "*The Data*", $\mathsf{L}, \mathsf{H}, \alpha, \beta$. However, fair-market values relate to \$0, so only scale-invariance is appropriate. We *could* force scale-invariance by dividing all the given data by, say, H, and then multiply the resulting random# by H, but we'll do it directly.]

Interval conditioning. We'll use "The Data" to produce a CDF F and density D, then condition them on interval $[\mathsf{L}, \mathsf{H}]$, producing $\widetilde{F}$ and $\widetilde{D}$. More precisely, our $\widetilde{F}$ will stretch vertically outside of $[0,1]$, but if we restrict our attention to $[\mathsf{L}, \mathsf{H}]$, then $\widetilde{F}$ <u>is</u> a CDF.

This *almost*-CDF corresponding to F is

2a:
$$\widetilde{F} := \mathcal{A} \circ F, \quad \textit{using affine} \quad \mathcal{A}(z) := \frac{z - \boldsymbol{\ell}}{\boldsymbol{m}}$$
$$\textit{with} \quad \boldsymbol{\ell} := F(\mathsf{L}) \quad \textit{and} \quad \boldsymbol{m} := F(\mathsf{H}) - F(\mathsf{L}). \quad \textit{So}$$
$$\widetilde{F}^{-1} := F^{-1} \circ \mathcal{A}^{-1}, \quad \textit{where} \quad \mathcal{A}^{-1}(y) := \boldsymbol{\ell} + \boldsymbol{m}y.$$

As $\widetilde{F}$ is a vertically stretched CDF, it can take on values >1 or <0. Nonetheless, Algorithm A applies to $\widetilde{F}^{-1}$, since the input is a point in $(0,1)$. The <u>actual</u> CDF is $C_{0,1} \circ \widetilde{F}$, using a ***cutoff function***

2b:
$$C_{\mathsf{Low},\mathsf{High}}(x) := \mathrm{Min}(\mathrm{Max}(x, \mathsf{Low}), \mathsf{High}),$$

for real numbers $\mathsf{Low} \le \mathsf{High}$.

When F is differentiable, so is $\widetilde{F}$.

2c: *Density $\widetilde{D}$ is derivative $\widetilde{F}'$, restricted to interval $[\mathsf{L}, \mathsf{H}]$.*

The only part of $\widetilde{D}$ we would wish to graph is on $[\mathsf{L}, \mathsf{H}]$. Thus we *never need* to work with $C_{0,1} \circ \widetilde{F}$ [the actual CDF] but only with $\widetilde{F}^{-1}$ and $\widetilde{F}'$.

Tools. The classic ***sigmoid*** fnc is

3a:
$$S(x) := \frac{\mathsf{e}^x}{\mathsf{e}^x + 1} \overset{\text{note}}{=\!=} \frac{1}{1 + \mathsf{e}^{-x}}, \quad \textit{with}$$
$$S^{-1}(y) = \log\Big(\frac{y}{1-y}\Big) \overset{\text{note}}{=\!=} \log\Big(\frac{1}{1-y} - 1\Big),$$

where $\log$ is natural logarithm, $\log_{\mathsf{e}}$. Our S is a strictly-incr map of $\mathbb{R}$ onto $(0,1)$. Under rotation by 180°, its graph is symmetric about the point $(0, \frac{1}{2})$.

For future reference, the sigmoid's derivative is

3b:
$$S'(x) = \frac{\mathsf{e}^x}{[\mathsf{e}^x + 1]^2}.$$

The rotational symmetry of $S()$ follows from showing that $S'()$ is even. Indeed,

$$S'(-x) = \frac{\mathsf{e}^{-x}}{[\mathsf{e}^{-x} + 1]^2} \cdot \Big[\frac{\mathsf{e}^x}{\mathsf{e}^x}\Big]^2 = \frac{\mathsf{e}^x}{[1 + \mathsf{e}^x]^2} \overset{\text{note}}{=\!=} S'(x).$$

Construction. Our CDF will have form $F(x) := S(\text{Multiplier} \cdot [x - \boldsymbol{\tau}])$, where ***translation*** $\boldsymbol{\tau}$ is the center of density D. [Center $\boldsymbol{\tau}$ might be outside of $[\mathsf{L}, \mathsf{H}]$, if the fair-market values mostly agree, and are well outside $[\mathsf{L}, \mathsf{H}]$.] To guarantee that the CDF is scale-invariant, the specific form will be

4a: $$F(x) := S\Big(\boldsymbol{c} \cdot \frac{x - \boldsymbol{\tau}}{\mathsf{L} + \mathsf{H}}\Big),$$

where ***concentration*** $\boldsymbol{c}$ is multiplicatively-invariant; multiplying L,H,α,β by a constant does not change $\boldsymbol{c}$. Increasing the positive number $\boldsymbol{c}$ will concentrate the probability around center $\boldsymbol{\tau}$.

A mult-invariant number is ***reliability***,

4b: $$\boldsymbol{\rho} := 1 - \frac{|\beta - \alpha|}{\beta + \alpha},$$

which is close to 1 when the buyer's and seller's fair-market opinions are relatively close.

Details. Varying with scale are aVerages

4c: $$V_{\mathsf{Fair}} := \frac{\alpha + \beta}{2} \quad \textit{and} \quad V_{\mathsf{Range}} := \frac{\mathsf{L} + \mathsf{H}}{2},$$

and weighted-average

4d: $$\boldsymbol{\tau} := V_{\mathsf{Fair}} \cdot \widehat{\boldsymbol{\rho}} + V_{\mathsf{Range}} \cdot [1 - \widehat{\boldsymbol{\rho}}],$$ *where* $\widehat{\boldsymbol{\rho}}$ *is some positive, non-decreasing function of* $\boldsymbol{\rho}$.

E.g, $\widehat{\boldsymbol{\rho}}$ could be $\boldsymbol{\rho}^3$.

Concentration $\boldsymbol{c}$ should also be a *positive, non-decreasing* fnc of $\boldsymbol{\rho}$. [Likely you want $\boldsymbol{c} \geq 1$ always. Isaac, we experimented with something like $\boldsymbol{c} := [0.2 + \boldsymbol{\rho}] \cdot 20$.]

Coding. Reals $M \neq 0$ and T determine an affine map

5: $$\mathcal{A}_{T,M}(z) := \frac{z - T}{M}, \quad \textit{with } \mathcal{A}^{-1}_{T,M}(y) := T + My.$$ *and (constant) derivative* $\mathcal{A}'_{T,M} = \frac{1}{M}$.

Decide your formulas for "dials" $\widehat{\boldsymbol{\rho}}$ and $\boldsymbol{c}$. Then:

Program. Define $\boldsymbol{\tau}$ by (4b,4c,4d), then let

6.1: $$F := S \circ \mathcal{A}_{\boldsymbol{\tau},M} \quad \textit{where} \quad M := \frac{\mathsf{L} + \mathsf{H}}{\boldsymbol{c}}.$$

Our *almost*-CDF is now

6.2: $$\widetilde{F} := \mathcal{A}_{\boldsymbol{\ell},\boldsymbol{m}} \circ F, \quad \textit{where}$$ $$\boldsymbol{\ell} := F(\mathsf{L}) \quad \textit{and} \quad \boldsymbol{m} := F(\mathsf{H}) - F(\mathsf{L}).$$

Algorithm A: Pick y uniformly in $(0,1)$. Using S^{-1} from (3a), compute $q := \mathcal{A}^{-1}_{\boldsymbol{\tau},M}\Big(S^{-1}(\mathcal{A}^{-1}_{\boldsymbol{\ell},\boldsymbol{m}}(y))\Big)$. I.e,

6.3: $$q := \boldsymbol{\tau} + M \cdot S^{-1}(\boldsymbol{\ell} + \boldsymbol{m}y)$$

is your random number. □

Graphing. Density $\widetilde{D}$ is the derivative of $\widetilde{F}$. I.e,

6.4: $$\widetilde{D}(x) = \frac{1}{\boldsymbol{m}} \cdot S'\Big(\frac{x - \boldsymbol{\tau}}{M}\Big) \cdot \frac{1}{M}$$

on $[\mathsf{L}, \mathsf{H}]$, using S' from (3b). □

NB. You *might* be able to avoid overflow[♡1] by using formulae $S(x) = \frac{1}{[1 + \mathrm{e}^{-x}]}$ and $S'(x) = \frac{\mathrm{e}^{-x}}{[\mathrm{e}^{-x} + 1]^2}$.

Alternatively, you can replace S by a sigmoid-like fnc, e.g

$$\mathfrak{S}(x) := \frac{1}{2} + \frac{\arctan(x)}{\boldsymbol{\pi}}. \quad \textit{Thus}$$

7: $$\mathfrak{S}^{-1}(y) = \tan([y - \tfrac{1}{2}] \cdot \boldsymbol{\pi}) \quad \textit{and}$$ $$\mathfrak{S}'(x) = \frac{1}{\boldsymbol{\pi} \cdot [1 + x^2]}.$$

[♡1] But you might have underflow. You can *guarantee* no overflow, by using that S' is <u>even</u>, hence $S'(x) = \frac{\mathrm{e}^{-|x|}}{[\mathrm{e}^{-|x|} + 1]^2}$.