The quantitative rules

Proof of 2.19

"We verify that $\partial V/\partial y = \partial U/\partial z$":

By definition $U = G(x, v)F_1(y, z)$, so by product rule and chain rule and using that $v = F(y, z)$ we have

$$U_z = G_2(x, v)F_2(y, z)F_1(y, z) + G(x, v)F_21(y, z).$$

Similarly $V = G(x, v)F_2(y, z)$ so

$$V_y = G_2(x, v)F_1(y, z)F_2(y, z) + G(x, v)F_12(y, z).$$

Of course mixed partials are equal (under some continuity of second derivatives assumption, Clairot's theorem), so the two expressions are the same.

Continuing to get 2.19 from 2.18:

We have then $G(x, y)G(y, z) = P(x, z)$. Pick any fixed z. Denote $P(x, z) = A(x)$ and $G(y, z) = B(y)$. Then $G(x, y) = \frac{A(x)}{B(y)}$ [and $G(y, z) = \frac{A(y)}{B(z)}$].

Plug this in to $G(x, y)G(y, z) = P(x, z)$ to get $\frac{A(x)A(y)}{B(y)B(z)} = P(x, z)$. So $A(y)/B(y)$ is independent of y, so is constant equal to r. This means

$G(x, y) = \frac{A(x)}{B(y)} = \frac{A(x)A(y)}{A(y)B(y)} = r\frac{A(x)}{A(y)}$

Proof of 2.23 and 2.24 from 2.22

The variables v, y, z are related by $v = F(y, z)$. One can interpret 2.22 as equality of differential 1-forms on the surface $\Sigma = \{v, y, z | v = F(y, z)\}$ in 3D space with coordinates $v - y - z$. The forms $\frac{v}{H(v)}$, $\frac{y}{H(y)}$ and $\frac{z}{H(z)}$ are exact, meaning are differentials of functions. We can find these functions by single-variable integration because each of the 1-forms depends only on one of the variables (in forms language, is pulled back via coordinate projection from the corresponding coordinate line), meaning we can find the antiderivative of the corresponding 1-form on the cordinate line, and then "pull back" the result, which just means interpret it as a function on Σ. Namely if we set

$$f(x) = \int_{x_0}^{x} \frac{1}{H(t)} dt$$

be a function well-defined up to a constant, we have

$$df(v) = \frac{dv}{H(v)}, \ df(y) = \frac{dy}{H(y)}, \ df(z) = \frac{dz}{H(z)}.$$

Then, since untiderivative of a 1-form on a connected Σ is well-defined up to a constant, equality of 1-forms

$$\frac{dv}{H(v)} = \frac{dy}{H(y)} + r\frac{dz}{H(z)}$$

implies equality of antiderivatives up to an additive constant, i.e.

$$f(v) = f(y) + rf(z).$$

Exponentiating this we get, on the surface $v = F(y, z)$ the equality, up to a multiplicative constant,

$$w(v) = w(y)w(z)^r$$

or

$$w(F(y, z)) = w(y)w(z)^r.$$

Brief explanation of the overall line of reasoning on from 2.45 to 2.58

TODO

Symmetry of the domain of 2.45

The domain is $0 \le S(y) \le x$ because $S(y)$ is plausibility of $\bar{B} = AD$, which is at most plausibility of A, aka x. for general A, D these (and $x, y \in [0, 1]$) are the only requirements, anything else should be possible, so hence the domain).

The symmetry of the domain comes from S being self-inverse and monotone decreasing. In fact, by monotonicity we have $S(y) \le x \Leftrightarrow S(S(y)) \ge S(x)$ and by $SS = Id$ we have $S(S(y)) \ge S(x) \Leftrightarrow y \ge S(x)$.

(In general the graph of the inverse function is obtained by flipping the graph of the original function through the $x = y$ line, and so the graph of S is symmetric to this flip preciely when S is it's own inverse; monotonicity makes the same true for the overgraph region.)

We can slightly rewrite the above argument as: $y \ge S(x)$ means $w(\overline{AD}) \ge w(\bar{A})$, which by monotonicity of S means $w(AD) \ge w(A)$, i.e. $x \ge S(y)$.

Proof of Equation 2.50

Source: stackexchange, I've reworded it and added detail to (hopefully) make it clearer.

We will use the Taylor series approximation, which is an approximation of $f(t)$ around the point a:

$$f(t) = f(a) + f'(a)(t - a) + O((t - a)^2)$$

Big O notation is described on Wikipedia.

The proof:

Letting $\delta = e^{-q}$, we have from (2.48):

$$S(y) = S\left[\frac{S(x)}{1 - \delta}\right]$$

We then use a Taylor series approximation of the function $f(\delta) = \frac{1}{1-\delta}$ around with $a = 0$.

$$S(y) = S[S(x)(1 + \delta + O(\delta^2))]$$

$$S(y) = S[S(x) + S(x)\delta + S(x)O(\delta^2)]$$

Now we want to get rid of the $S[]$ surrounding the equation, so we will use another Taylor approximation of the function $S(t)$. We approximate around the point $a = S(x)$.

This gives us the approximation of $S(t)$ as:

$$S(t) = S[S(x)] + S'[S(x)](t - S(x)) + O((t - S(x))^2)$$

Letting $t = S(x) + S(x)\delta + S(x)O(\delta^2)$

$$S[S(x)+S(x)\delta+S(x)O(\delta^2)] = S[S(x)]+S'[S(x)](S(x)\delta+S(x)O(\delta^2))+O((S(x)\delta+S(x)O(\delta^2))^2)$$

$$S[S(x)+S(x)\delta+S(x)O(\delta^2)] = S[S(x)]+S'[S(x)]S(x)\delta+S'[S(x)]S(x)O(\delta^2)+O((S(x)\delta+S(x)O(\delta^2))^2)$$

With big O notation we can get rid of constant factors:

$$S[S(x)+S(x)\delta+S(x)O(\delta^2)] = S[S(x)]+S'[S(x)]S(x)\delta+O(\delta^2)+O((\delta+O(\delta^2))^2)$$

With big O notation we can also get rid terms that drop asymptotically faster than the largest term.

$$S[S(x) + S(x)\delta + S(x)O(\delta^2)] = S[S(x)] + S'[S(x)]S(x)\delta + O(\delta^2)$$

□

Explanation of 2.52, 2.53 (incomplete)

2.45 says $S[S(x)] = x$. Differentiating in x we get $S'[S(x)]S'(x) = 1$, or $S'[S(x)] = 1/S'(x)$. Now we plug in into 2.50 to get

$$S(y) = x + \exp\{-q\}S(x)/S'(x) + O(\exp\{-2q\})$$

Denoting by $\alpha(x) = \log\left[\frac{-xS'(x)}{S(x)}\right]$ we get

$$S(y) = x + \exp\{-q\}(-x\exp-\alpha) + O(\exp\{-2q\})$$

Dividing by x

$$\frac{S(y)}{x} = 1 - \exp\{-(q+\alpha)\} + \frac{1}{x}O(\exp\{-2q\})$$

which is a version of 2.51.

From now on we will treat x as fixed and only vary q, sening it to $+\infty$, which in light of 2.48 means keeping x fixed and sending y to $S(x)$ from below.

Then we can write

$$\frac{S(y)}{x} = 1 - \exp\{-(q+\alpha)\} + O(\exp\{-2q\}),$$

which is 2.51.

Now we want to deduce 2.53. We make some progress but ultimately do not succeed (yet).

We start with 2.45

$$xS\left[\frac{S(y)}{x}\right] = yS\left[\frac{S(x)}{y}\right]$$

and plug in 2.51 and 2.48 to get

$$xS[1 - \exp\{-(q+\alpha)\} + O(\exp\{-2q\})] = yS[1 - \exp\{-q\}]$$

Right hand side is $y\exp\{-J(q)\}$ by definition 2.49. We also plug in 2.48 in the form $y = S(x)/(1 - \exp\{-q\})$ to get

$$RHS = S(x)\exp\{-J(q)\}/(1 - \exp\{-q\})$$

Now take log to get

$$\log x + \log S[1 - \exp\{-(q+\alpha)\} + O(\exp\{-2q\})]$$

$$= \log S(x) - J(q) - \log(1 - \exp\{q\})$$

Now if we could write

$$\log S[1 - \exp\{-(q+\alpha)\} + O(\exp\{-2q\})] =$$
$$\log S[1 - \exp\{-(q+\alpha)\}] + O(\exp\{-2q\})$$

we would get $J(q+\alpha) + O(\exp\{-2q\})$ and 2.53 would follow.

TODO figure out what to do about this

Proof of 2.57

The difficulty is getting 2.54 from 2.53. More precisely, we only need that

$$b = \alpha^{-1}\log\left[\frac{x}{S(x)}\right] = \log\left[\frac{x}{S(x)}\right] / \log\left[\frac{-xS'(x)}{S(x)}\right]$$

is constant. After that it's algebra:

$$b\log\left[\frac{-xS'(x)}{S(x)}\right] = \log\left[\frac{x}{S(x)}\right]$$

and exponentiating one gets

$$\left[\frac{-xS'(x)}{S(x)}\right]^b = \left[\frac{x}{S(x)}\right]$$

or

$$\left[\frac{-xS'(x)}{S(x)}\right] = \left[\frac{x}{S(x)}\right]^{1/b}$$

$$S'(x) = -S(x)^{1-1/b}x^{1/b-1}$$

$$\frac{dS}{dx} = -\frac{x^{m-1}}{S^{m-1}}$$

for $m = 1/b$.

I find it simpler to show that $b(x)$ is constant than to show the asymptotic expansion 2.54.

Remark: To get 2.54 one must first make sure that α actually takes "continuum of values". Since α is a continuous function of x, intermediate value theorem implies that the set of values of α is an interval; we just check that it is not a degenerate interval consisting of a single point. Indeed, that would mean $\alpha(x)$ is constant, or $S'(x)/S(x) = -c/x$, $(\ln S)' = -c/x$, $\ln S(x) = a - c/x$, but this breaks $S(0) = 1$. Now we do know that α takes ''continuum of values".

We start with 2.53 in the form

$$J(q + \alpha(x)) - J(q) = \beta(x) + O(\exp\{-q\})$$

We want to deduce that

$b(x) = \beta(x)/\alpha(x)$ is constant.

Intuitively, $J(q + \alpha(x)) - J(q) = \beta(x) + O(\exp\{-q\})$ does say that for every increment of $\alpha(x)$ in the input, the output of J increases by $\beta(x)$ (plus a small error), so (asymptotically) J must be linear with slope=rise/run=$b(x)$, and since there can be only one slope, $b(x)$ must be constant. The question is how to make it precise.

First, if the error term was absent

$$J(q + \alpha(x)) - J(q) = \beta(x)$$

implies that if $\alpha(x_1) = \alpha(x_2)$ then $\beta(x_1) = \beta(x_2)$, so β is a well-defined function of α, and if $J(q)$ is continuous $\beta(\alpha)$ is also continuous. Now we can write

6

$$J(q + \alpha) - J(q) = \beta(\alpha).$$

Now this implies by induction

$$J(q + n\alpha) = J(q) + n\beta(\alpha)$$

Then given any two α values α_0, α_1, we have

$$J(q + n_0\alpha_0) = J(q) + n_0\beta(\alpha_0),$$

$$J(q + n_1\alpha_1) = J(q) + n_1\beta(\alpha_1).$$

If α_0/α_1 is rational then $\alpha_1 = (n_0/n_1)\alpha_0$, and after plugging into the above

$$J(q + n_0\alpha_0) = J(q) + n_0\beta(\alpha_0) = J(q) + n_1\beta(\alpha_1)$$

so

$$\beta(\alpha_0)/\beta(\alpha_1) = n_0/n_1 = \alpha_0/\alpha_1$$

meaning $\beta(\alpha_0)/\alpha_0 = \beta(\alpha_1)/\alpha_1$. Since $\beta(\alpha)$ is continuous, bing constant on all rational multiples of a given α_0 implies that it is constant (recall α varies over an interval, on which rational multiples of α_0 are dense).

Now we want to repeat this argument with error terms.

1. Suppose x_1 and x_2 are such that $\alpha(x_1) = \alpha(x_2)$.

   Then $J(q + \alpha(x)) - J(q) = \beta(x) + O(\exp\{-q\})$ implies, by plugging in sufficiently large q, that the difference $\beta(x_1) - \beta(x_2)$ is smaller than any positive number, so is zero. Thus, as before $\beta(\alpha)$ is well-defined.

   To see that $\beta(\alpha)$ is continuous, given x_0 corresponding to some α_0 and any $\varepsilon > 0$ pick q such that $|O(\exp\{-q\})| < \varepsilon/2$ and, using continuity of J at $q + \alpha_0$, pick δ such that $|\alpha - \alpha_0| < \delta$ implies $|J(q + \alpha) - J(q + \alpha_0)| < \varepsilon/2$. Then $|\beta(\alpha) - \beta(\alpha_0)| < \varepsilon$ on the same interval $|\alpha - \alpha_0| < \delta$, meaning that $\beta(\alpha)$ is continuous at α_0, as wanted.

2. We know that for each x and each $C > 0$ there exists $Q(x, C)$ such that that for $q \geq Q(x, C)$ we have $|O(\exp\{-q\})| < C \exp\{-q\}$. Pick any $q(x, C) \geq Q(x, C)$.

Now we have by induction (with everything depending on x)

$$|J(q+n\alpha)-(J(q)+n\beta)|$$

$$\leq C \exp\{-q\}(1+\exp\{-\alpha\}+...+\exp\{-(n-1)\alpha\})$$

$$<\frac{C}{1-\exp\{-\alpha\}} \exp\{-q\}$$
As before is $\alpha_1=(n_0/n_1)\alpha_0$ writing the above and picking sufficiently

$$\beta(\alpha_0)/\beta(\alpha_1)=n_0/n_1=\alpha_0/\alpha_1.$$
The rest is the sam as in the "error-less" case.

This shows that

$$b(x) = \beta x/\alpha x = \log\left[\frac{x}{S(x)}\right] / \log\left[\frac{-xS'(x)}{S(x)}\right]$$

is constant, and thus establishing 2.57.

Proof of 2.58

$S^{m-1}S' = -x^{m-1}$ is equivalent to $(S^m)' = -\frac{1}{m}x^{m-1}$, so that $S^m = C - x^m$.
Initial value $S(0) = 1$ fixes $C = 1$ and $S(x) = (1 - x^m)^{1/m}$ as wanted.

Exercise 2.1

I think this problem is ambiguous and can be interpreted in multiple ways, see here for a different interpretation. But I think the following interpretation makes more sense.

With X representing any background information:

$$p(C|(A + B)X) = \frac{p(A + B|CX)p(C|X)}{p(A + B|X)}$$

$$= \frac{(p(A|CX) + p(B|CX) - p(AB|CX))p(C|X)}{p(A|X) + p(B|X) - p(AB|X)}$$

$$= \frac{p(AC|X) + p(BC|X) - p(ABC|X)}{p(A|X) + p(B|X) - p(AB|X)}$$

Exercise 2.2

We will use convention that **all P are conditioned on** X. So $P(A|C)$ actually stands for $P(A|CX)$.

First we do a bunch of lemmas about mutually exclusive propositions.

1. If A_i and are mutually exclusive, and C is arbitrary, then

   a) $P(A_i + A_j) = P(A_i) + P(A_j)$

   Proof: $P(A_i + A_j) = P(A_i) + P(A_j) - P(A_i A_j) = P(A_i) + P(A_j)$.

   b) $A_i C$ are mutually exclusive

   Proof: If $i \neq j$ then $P(A_i C A_j C) = P(A_i A_j) P(C|A_i A_j) = 0$.

   c) $A_i|C$ are mutually exclusive

   Proof: If $i \neq j$ then $P(A_i|C)P(A_j|C) = P(A_i C)P(A_j C)/P(C)^2 = 0$.

2. If A_1, A_2, A_3 are mutually exclusive, then $A_1 + A_2$ and A_3 are mutually exclusive.

   First of all $P(A_1 A_2 A_3) = P(A_1|A_2 A_3)P(A_2 A_3) = 0$. Then,

$$P((A_1 + A_2)A_3) = P(A_1 A_3 + A_2 A_3) =$$

$$P(A_1 A_2) + P(A_2 A_3) - P(A_1 A_2 A_3) = 0.$$

With this in place, we can use induction to see

$$P\left(\sum A_i\right) = \sum P(A_i)$$

and

$$P\left(C\left(\sum A_i\right)\right) = \sum P(CA_i).$$

Finally,

$$P\left(C\left(\sum A_i\right)\right) = P\left(C\middle|\left(\sum A_i\right)\right)P\left(\sum A_i\right)$$

and plugging in we get

$$P\left(C\middle|\left(\sum A_i\right)\right) = \frac{P(C(\sum A_i))}{P(\sum A_i)} = \frac{\sum P(CA_i)}{\sum P(A_i)} = \frac{\sum P(A_i)P(C|A_i)}{\sum P(A_i)}.$$

Exercise 2.3

Again, everything is conditional on C, but we don't write it.

Then

$P(AB) = P(B|A)P(A) \leq P(A) = a,$

$P(A + B) = P(A) + P(B) - P(AB) = a + b - P(AB) \geq b.$

Also

$P(AB) = P(A) + P(B) - P(A + B) \geq a + b - 1$

and

$P(A + B) = P(A) + P(B) - P(AB) \leq a + b.$