The quantitative rules

Proof of 2.19

Credit to atwwgb

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

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

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

TODO

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.

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

$\square$

Explanation of 2.52, 2.53

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, sending 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.

Proof of 2.57

Start with 2.56 and doing a bit of manipulation to isolate

$$S'(x)$$

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

Expanding $S'(x)$ into the actual derivative, and treating them as differentials.

$$\begin{aligned}
\frac{dS(x)}{dx} &= -x^{\frac{1}{b}-1}S(x)^{1-\frac{1}{b}} \\
S(x)^{\frac{1}{b}-1}dS &= -x^{\frac{1}{b}-1}dx \\
S(x)^{\frac{1}{b}-1}dS + x^{\frac{1}{b}-1}dx &= 0 \\
S(x)^{m-1}dS + x^{m-1}dx &= 0
\end{aligned}$$

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

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

and

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

Finally,

$$P(C(\sum A_i)) = P(C|(\sum A_i))P(\sum A_i)$$

and plugging in we get

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

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

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