Dilina's Algorithm − Unreliable Means and Variances

Given a new user, we want to compute $P(G = g|D(t), x_i)$ where $D(t)$ is the history of the new user on turn t and x_i are some observations gathered before we start the algorithm. In particular x_i are some draws from each group for each item, used to estimate group means and variances.

If we know the means and variances a priori, we can compute $P(D(t)|G = g)$ exactly. However, if we have few observations x_i then we cannot be confident the sample mean and variances are close to the true means and variances, and we should be more conservative updating the probabilities.

Write $\mathbf{\Sigma} = \sigma(g, v)_{gv}$ and $\mathbf{M} = \mu_{gv}$ for the matrices of means and standard deviations. First decompose over all the possible σ, μ values on an evenly-spaced grid:

$$P(G = g|D, x) = \sum_{\sigma,\mu} P(G = g, \mathbf{M} = \mu, \mathbf{\Sigma} = \sigma|D, x) \tag{1}$$

$$= \sum_{\sigma,\mu} P(g, \mu, \sigma|D, x_i) \text{ for ease of notation} \tag{2}$$

where we sum over all possibilities for $\mathbf{M}, \mathbf{\Sigma}$. Using Bayes' theorem we can flip the two arguments on the RHS to get

$$P(g, \mu, \sigma|D, x) = P(D, x|g, \mu, \sigma)\frac{P(g, \mu, \sigma)}{P(D, x)} \tag{3}$$

Since x are drawn before we start the algorithm, and we assume all the group-item draws are independent, the first term factors as $P(D|g, \mu, \sigma)P(x|g, \mu, \sigma)$. The first factor, we already know how to compute it under the normal assumption.

$$P(D|G = g, \mu, \sigma) = \prod_{i=1}^{t} \frac{1}{\sigma(g, v_i)\sqrt{2\pi}} \exp\left(-\frac{(r_i - \mu(g, v_i))^2}{\sigma^2(g, v_i)}\right)$$

The above must be re-computed each turn when D is updated. But the exponential form means we just need to multiply the old RHS by a new exponential factor

For the second factor, again use how x are drawn before we start the algorithm, to eliminate the dependence on g from the conditional expectation, to get $P(x|g, \mu, \sigma) = P(x|\mu, \sigma)$.

Since all the draws are independent, this factorises as the product of drawing each x individually. To write it in full, say we make $i(g, v)$ draws each from group g and item v. So $x = \{x(g, v, i) : i \leq i(g, v)\}$.

$$P(x|\mu, \sigma) = \prod_{i,g,v} \frac{1}{\sigma(g,v)\sqrt{2\pi}} \exp\left(-\frac{(x(i,g,v) - \mu(g,v))^2}{\sigma^2(g,v)}\right)$$

$$= \prod_{g,v} \prod_{i=1}^{i(g,v)} \frac{1}{\sigma(g,v)\sqrt{2\pi}} \exp\left(-\frac{(x(i,g,v) - \mu(g,v))^2}{\sigma^2(g,v)}\right)$$

$$= \prod_{g,v} \left(\frac{1}{\sigma(g,v)\sqrt{2\pi}}\right)^{i(g,v)} \exp\left(-\sum_{i=1}^{i(g,v)} \frac{(x(i,g,v) - \mu(g,v))^2}{\sigma^2(g,v)}\right)$$

Going back to (2) and (3) we get

$$P(G = g|D, x) = \sum_{\mu,\sigma} P(x|\mu,\sigma)P(D|g,\mu,\sigma)\frac{P(g,\mu,\sigma)}{P(D,x)} \tag{4}$$

where the first two factors are known. For the third factor use independence to write the numerator as $P(g,\mu,\sigma) = P(g)P(\mu,\sigma)$. The denominator does not depend on g and so we can neglect it to get the final form

$$P(G = g|D, x) \propto \sum_{\mu,\sigma} P(x|\mu,\sigma)P(D|g,\mu,\sigma)P(\mu,\sigma)P(g) \tag{5}$$

The first factor $P(x|\mu, \sigma)$ depends on the samples taken to estimate the clusters. It does not change each turn. The second factor $P(D|g, \mu, \sigma)$ must be updated with each new item presented. The $P(\mu, \sigma), P(g)$ will depend on priors for g and μ, σ. The former can be taken as the proportion of users in each cluster. The latter will have to be a "measure of uncertainty". We cannot take it as the empirical distribution of all the arms, since the distinguisher arms are by definition non-representative of the other arms as a whole. I suggest a uniform prior on μ, σ. Or perhaps uniform priors on μ, σ^2 since both the variance and standard deviations occur in the formulae.

Let's do a sanity check to recover the formula from earlier, where we assumed μ, σ are known a priori. Notice we only used the identity of x when we (a) compute $P(x|\mu, \sigma)$ explicitly and (b) write $P(D, x|g, \mu, \sigma) = P(D|g, \mu, \sigma)P(x|g, \mu, \sigma)$. So the formula (5) holds for any random variable x such that (b) holds. In particular let $x = x_0$ be constant and $\Sigma = \sigma_0, \mathbf{M} = \mu_0$ also be constant. Since constant random variables are independent of everything else we still have (b) and the formulas will simplify.

The LHS of (5) becomes just $P(G = g|D)$. For the RHS the sum disappears since the term $P(\mu, \sigma) = P(\mathbf{M} = \mu, \mathbf{\Sigma} = \sigma)$ is zero except when $\mu, \sigma = \mu_0, \sigma_0$ and it equals 1.

$$P(G = g|D) \propto P(x_0|\mu_0, \sigma_0)P(D|g, \mu_0, \sigma_0)P(g) \tag{6}$$

The first term is constant and cancels when we normalise. probability 1 the above becomes

$$P(G = g|D) \propto \frac{P(D|g, \mu_0, \sigma_0)P(g)}{\sum_h P(D|h, \mu_0, \sigma_0)P(h)} \tag{7}$$

which is the same as before.

0.1 But what does it all mean?

The meaning of a prior on G is clear. It says what is the chance of a randomly drawn user coming from a given group. This can be estimated in practice by looking at past users. Then Dilina's algorithm gives a probability as we draw a large number of users. A uniform prior means the following: Fix a history D and generate a large number of users with equal probability from each group. Look at all the users with history D. The formula $\gamma^t(g)e^{-L^t(g)}$ gives what proportion of those users are in each group.

Having a prior over the user-items is slipperier. If we had a large sample of "useful items" to form a prior, we would already be done, since ther eis no need for all this is we have reliable samples. I think we will end up taking a uniform prior on the mean in some range, and then a uniform prior on either the standard deviation or variance in some other range. These will give different updates.

0.2 Uniform Priors

If we use a uniform prior on μ, σ, g the $P(\mu, \sigma)$ and $P(g)$ terms cancel and we get

$$P(G = g|D, x) \propto \sum_{\mu, \sigma} P(x|\mu, \sigma)P(D|g, \mu, \sigma) \tag{8}$$

0.3 Computing the Parts

The first factor we want to compute is

$$P(x|\mu,\sigma) = \prod_v \prod_h \left(\frac{1}{\sigma(h,v)}\right)^{i(h,v)} \exp\left(-\sum_{i=1}^{i(h,v)} \frac{(x(i,h,v) - \mu(h,v))^2}{\sigma^2(h,v)}\right)$$

$$= \prod_v \prod_h X_{hv} = \prod_v \left(X_{gv} \prod_{h \neq g} X_{hv}\right)$$

for any g. Note since we only interested in the proportions of (5) as g varies we have neglected the factors of $\sqrt{2\pi}$ above. We can also factorise

$$P(D|G = g, \mu, \sigma) = \prod_{i=1}^{t} \frac{1}{\sigma(g, v_i)} \exp\left(-\frac{(r_i - \mu(g, v_i))^2}{\sigma^2(g, v_i)}\right)$$

To make this amenable to products with $P(x|\mu,\sigma)$ write, on turn t, for each v that $\mu(g, \widetilde{v}) = \mu(g, v_i)$ if there is some turn $i = t, \ldots, t$ with $v_i = v$. Since we never present the same item twice all v_i are distinct and so $\widetilde{v}$ is well-defined. Likewise define $r(\widetilde{v}) = r_i$ for the same turn i. i.e $r(\widetilde{v})$ is the reward on the turn v was presented to the user. Define $\sigma(g, \widetilde{v}) = \sigma(g, v_i)$ for the same i. If no such i exists define $r(\widetilde{v}) = \mu(g, \widetilde{v}) = 0$ and $\sigma(g, \widetilde{v}) = 1$. The above can be written

$$P(D|G = g, \mu, \sigma) = \prod_v \frac{1}{\sigma(g, \widetilde{v})} \exp\left(-\frac{(r(\widetilde{v}) - \mu(g, \widetilde{v}))^2}{\sigma^2(g, \widetilde{v})}\right) = \prod_v Y_{gv}$$

since each v-factor equals 1 unless v has been played. Hence we can write

$$P(G = g|D, x) \propto \sum_{\mu,\sigma} \left(\prod_v Y_g\right) \prod_v \left(X_{gv} \prod_{h \neq g} X_{hv}\right) = \sum_{\mu,\sigma} \prod_v \left(X_{gv} Y_{gv} \prod_{h \neq g} X_{hv}\right) \quad (9)$$

Unfortunately, integration over σ, μ involves integrating simultaneously over each $\mu(g, v)$ and $\sigma(g, v)$ of which there are likely hundreds. i.e $\sum_{\mu,\sigma}$ abbreviates

$$\sum_{\substack{\mu(G,I) \\ \sigma(G,I)}} \sum_{\substack{\mu(G-1,I) \\ \sigma(G-1,I)}} \cdots \sum_{\substack{\mu(1,2) \\ \sigma(1,2)}} \sum_{\substack{\mu(G,1) \\ \sigma(G,1)}} \sum_{\substack{\mu(G-1,1) \\ \sigma(G-1,1)}} \cdots \sum_{\substack{\mu(1,1) \\ \sigma(1,1)}}$$

Even assuming independence and separable priors, we still end up integrating over a box with dimension in the hundreds or thousands. Fortunately things decompose nicely to let

us only integrate over relatively smaller boxes, one corresponding to each g, v pair. Observe that $\mu(f, v)$ and $\sigma(f, v)$ appear only in the terms $X_{f,v}$ and $Y_{f,v}$ and not in the other $X_{f',v'}$ and $Y_{f',v'}$ Hence we can bring the sums through the products to get

$$P(G = g|D, x) \propto \prod_v \left(\left(\sum_{\substack{\mu(g,v) \\ \sigma(g,v)}} X_{gv} Y_{gv} \right) \prod_{h \neq g} \left(\sum_{\substack{\mu(h,v) \\ \sigma(h,v)}} X_{hv} \right) \right) \tag{10}$$

$$= \left(\prod_v \sum_{\substack{\mu(g,v) \\ \sigma(g,v)}} X_{gv} Y_{gv} \right) \left(\prod_v \prod_{h \neq g} \sum_{\substack{\mu(h,v) \\ \sigma(h,v)}} X_{hv} \right) \tag{11}$$

The second term depends only on the data obtained before we run the algorithm. The first term changes as we present more items and as more $\widetilde{v}$ become active. We can simplify somewhat by rescaling by $\prod_v \prod_h \sum_{\substack{\mu(h,v) \\ \sigma(h,v)}} X_{hv}$ to get

$$P(G = g|D, x) \propto \left(\prod_v \sum_{\substack{\mu(g,v) \\ \sigma(g,v)}} X_{gv} Y_{gv} \right) / \left(\prod_v \sum_{\substack{\mu(g,v) \\ \sigma(g,v)}} X_{gv} \right) \tag{12}$$

The denominator depends only on g and does not change as the history changes. Now reintroduce the dependence on time through D^t and Y_{gv}^t. We can obtain the above on turn t from the equivalent on turn $t - 1$ as follows

$$P(G = g|D^t, x) \propto P(G = g|D^{t-1}, x) \times \left(\prod_v \sum_{\substack{\mu(g,v) \\ \sigma(g,v)}} X_{gv} Y_{gv}^t \right) / \left(\prod_v \sum_{\substack{\mu(g,v) \\ \sigma(g,v)}} X_{gv} Y_{gv}^{t-1} \right) \tag{13}$$

According to the definition the values of Y_{gv}^{t-1} and Y_{gv}^t change only for $v = v(t)$ the action on turn t. Also since $v(t)$ has not been played before turn t we have $Y_{gv(t)}^t = 1$ and

$$Y_{gv(t)}^t = \frac{1}{\sigma(g, v(t))} \exp \left(-\frac{(r_t - \mu(g, v(t)))^2}{\sigma^2(g, v(t))} \right)$$

Hence the other factors in the v-product cancel and we get

$$P(G = g|D^t, x) \propto P(G = g|D^{t-1}, x) \times \left(\sum_{\substack{\mu(g,v(t)) \\ \sigma(g,v(t))}} X_{gv(t)} Y^t_{gv(t)} \right) \Big/ \left(\sum_{\substack{\mu(g,v(t)) \\ \sigma(g,v(t))}} X_{gv(t)} \right) \quad (14)$$

We can now recall the definitions of X_{gv} and Y_{gv} to expand the rightmost factor as

$$\frac{\sum \left(\frac{1}{\sigma(g,v(t))} \right)^{i(g,v(t))+1} \exp \left(-\sum_{i=1}^{i(g,v(t))} \frac{(x(i,g,v)-\mu(g,v(t)))^2}{\sigma^2(g,v(t))} - \frac{(r_t-\mu(g,v(t)))^2}{\sigma^2(g,v(t))} \right)}{\sum \left(\frac{1}{\sigma(g,v(t))} \right)^{i(g,v(t))} \exp \left(-\sum_{i=1}^{i(g,v(t))} \frac{(x(i,g,v)-\mu(g,v(t)))^2}{\sigma^2(g,v(t))} \right)} \quad (15)$$

or supressing the indices on μ and σ where possible

$$\frac{\sum_{\mu,\sigma} \sigma^{-i(g,v(t))-1} \exp \left(-\sum_{i=1}^{i(g,v(t))} \frac{(x(i,g,v)-\mu)^2}{\sigma^2} - \frac{(r_t-\mu)^2}{\sigma^2} \right)}{\sum_{\mu,\sigma} \sigma^{-i(g,v(t))} \exp \left(-\sum_{i=1}^{i(g,v(t))} \frac{(x(i,g,v)-\mu)^2}{\sigma^2} \right)} \quad (16)$$

This means each time we update the prior, we have to compute one summation per group.

Note: It feels like this can be derived in a more straightforward manner once you know what you're looking for.

Each sum should be computed numerically by replacing it with an integral, since we are summing over a fine grid in the relevant range. For example μ on a grid over $[0, 5]$ and σ over $[0, 2]$.

Note: There must be a way to make this formal, and have an integral rather than a sum from the very start. You would need a continuous version of the equality $P(A) = \sum_i P(A, Y = y_i)$ for Y a discrete random variable. Though in practice all our random variables are discrete, since we only have finite precision.

Swapping between sums and integrals will introduce a constant factor depending on the grid. But it cancels in the numerator and denominator provided we replace the top and bottom sums with integrals.

This should also work for a nonuniform prior on μ, σ provided they are separable. i.e $P(\mu, \sigma)$ factors as the product of all the $P(\mu(g, v))P(\sigma(g, v))$ terms. You just have to carry around the $P(\mu, \sigma)$ in the derivation, rather than setting it equal to 1, until you're ready to replace the sums with integrals. Instead we get

$$\frac{\sum_{\mu,\sigma} \sigma^{-i(g,v(t))-1} \exp\left(-\sum_{i=1}^{i(g,v(t))} \frac{(x(i,g,v)-\mu)^2}{\sigma^2} - \frac{(r_t-\mu)^2}{\sigma^2}\right) P(\Sigma(g,v(t)) = \sigma)P(M(g,v(t)) = \mu))}{\sum_{\mu,\sigma} \sigma^{-i(g,v(t))} \exp\left(-\sum_{i=1}^{i(g,v(t))} \frac{(x(i,g,v)-\mu)^2}{\sigma^2}\right) P(\Sigma(g,v(t)) = \sigma)P(M(g,v(t)) = \mu)}$$

$$(17)$$

For a fine grid the numerator becomes

$$\int_\mu \int_\sigma \sigma^{-i(g,v_t)-1} \exp\left(-\sum_{i=1}^{i(g,v_t)} \frac{(x(i,g,v)-\mu)^2}{\sigma^2} - \frac{(r_t-\mu)^2}{\sigma^2}\right) f_\Sigma(\sigma) f_M(\mu) d\mu d\sigma \qquad (18)$$

Where $f_\Sigma(\sigma), f_M(\mu)$ are the pdf of the random variables $\Sigma = \Sigma(g, v_t)$ and $M = M(g, v_t)$.

One prior that might be of interest is when the mean is still uniform, meaning $f_M(\sigma) = 1$ but, rather than the standard deviations being uniform, the variances are uniform, i.e $f_{\Sigma^2}(\sigma) = 1$. Write $\Sigma^2 = u(\Sigma)$ for $u(x) = x^2$. The general rule for pdfs of functions of random variables is

$$f_{u(\Sigma)}(\sigma) = u'(\sigma) f_\Sigma(u(\sigma))$$

assuming Σ^2 is uniform we get $1 = f_{\Sigma^2}(\sigma) = f_{u(\Sigma)}(\sigma) = u'(\sigma) f_\Sigma(u(\sigma)) = 2\sigma f_\Sigma(\sigma^2)$. Replace σ with σ^2 to get $f_\Sigma(\sigma) = 1/2\sqrt{\sigma}$. (**Note:** Double check this is the correct rule and usage, and the factor is not upside-down.)

There is a temptation to try and integrate symbolically. Integrating over σ first gives and answer in terms of the incomplete gamma function, which we then have to integrate over μ. Integrating over μ is an incomplete Gaussian integral, which we know has no closed form. However it certainly has lookup tables, so it might be faster to compute the integral in terms of the error function first, and then sum numerically over σ.

Perhaps it would not effect performance much if we replace the improper Gaussian integrals with proper Gaussian integrals. We know how to compute those.

0.4   UCB

The process for updating the test prior during the UCB step is the same as (16) with one difference − we cannot draw a test value for r_t since we do not know the mean and variance. The best we can do is select a test action w and draw the test rating r_t empirically from the samples $x(i, g, w)$ for group g on item w. The test prior values $P(G = g|\widetilde{D}^t, x)$ are obtained by multiplying $P(G = g|D^{t-1}, x)$ by

$$\frac{\int_{\mu,\sigma} \sigma^{-i(g,w)-1} \exp\left(-\sum_{i=1}^{i(g,w)} \frac{(x(i,g,v)-\mu)^2}{\sigma^2} - \frac{(S_{g,v}-\mu)^2}{\sigma^2}\right) d\Sigma dM}{\int_{\mu,\sigma} \sigma^{-i(g,w)} \exp\left(-\sum_{i=1}^{i(g,w)} \frac{(x(i,g,v)-\mu)^2}{\sigma^2}\right) d\Sigma dM} \tag{19}$$

where $S_{g,v} = x(i,g,v)$ with equal probability over i.

Question How does the above reduce to the old update when there is a large number of samples? Looking at it, it looks like the new term will be swallowed up by all the existing samples. What we'd expect to happen is that for a large number of samples, a lot of weight is put on a very small range of μ, σ, over that range there is very little variation, in the common factors of numerators and denominator, and they effectively cancel, leaving just the new sample factor, which matches Dilina's old algorithm

I suspect the inability to draw test samples is a blessing in disguise, as it will lower the computational cost of pulling an arm each UCB round. And we want to avoid computing double integrals in each UCB step since we might need 50,000 rounds for eight cluster problems. That's eight double integrals times 50,000 times however many real items we present to the user. Sounds like a lot.

For example if we have only five samples per group-item, then for each w, g there are only five values the factors (19) can take. So at the very start of the algorithm we compute the $G \times \mathcal{V} \times 5$ integrals offline once and for all, and then just look up the values for every UCB step. This works because the update factors do not change depending on the past history. The history for ever turn is incorporated into the prior on the following turn. From that point on we are free to forget the history.

References