Final Fall 2017 Solution

Noah Arthurs

June 5, 2019

Disclaimer: These solutions are not super verbose or complete. They're just meant to give you an idea of how to do each problem. Feel free to start discussions on Piazza if you're confused.

1. for a-d, see section 2 problem 4. For e:

$$P(A) = P(A|M)P(M)+P(A|M^C)P(M^C) = P(A|M^C)(P(M)+P(M^C)) = P(A|M^C)$$

2. (a) $P(X > 4) = 1 - P(X < 4) = 1 - F_X(4) = 1 - (1 - (1 + 4)^{-2}) = 1/25$

   (b) $E[X] = 0.5$, $Var(X) = 0.75$

   Let Y be sum of samples. Therefore by CLT, $Y \sim N(50, 75)$.

   $P(Y > 55) = 1 - P(Y < 55) = 1 - \Phi((55 - 50)/\sqrt{75})$

   (c) Assume each x_i is drawn from $X \sim ExamPareto(\alpha)$:

   $$LL(\alpha) = \log \Pi_{i=1}^{n} f_X(x_i) = \Sigma_{i=1}^{n} \log(\alpha(1+x_i)^{-(\alpha+1)}) = \Sigma_{i=1}^{n}(\log(\alpha) - (\alpha+1)\log(1+x_i))$$

   (d) Gradient ascent would be good (we have a 1-d gradient):

   $\nabla_\alpha LL(\alpha) = \frac{dLL}{d\alpha} = \Sigma_{i=1}^{n}(1/\alpha - \log(1 + x_i)) = n/\alpha - \Sigma_{i=1}^{n} \log(1 + x_i)$

   Even better is to find the point of inflection directly. We do this by setting $\frac{dLL}{d\alpha} = 0$:

   $$n/\alpha - \Sigma_{i=1}^{n} \log(1 + x_i) = 0 \implies \alpha = \frac{n}{\Sigma_{i=1}^{n} \log(1 + x_i)}$$

   To double check that this is a maximum:

   $$\frac{d^2\alpha}{d\alpha^2} = \frac{-n}{\alpha^2}$$

   Since n is positive, this is negative, so we've found our maximum!

(a) Number of passwords of length k is given by 26^k.

Number of passwords of length 5-10 is given by $\Sigma_{k=5}^{10} 26^k$

(b) See solution to Section 5 Problem 3 Part a.

(c) See solution to Section 5 Problem 3 Part b.

(d) I suspect that the answer they were looking for was p^5, as you need each of 5 incorrect-length (4 line) runs to be slower than the correct-length (6 line) run. However, this is not right! Because there is only one length of the correct run, the incorrect run lengths are dependent! As an intuition, if 4 out of 5 incorrect runs are slower than the correct run, that would make us think that the correct run was unusually fast and thus believe that the 5th incorrect run has a good chance of being slower as well. So let's do it better!

We need the probability that each of 5 incorrect-length runs is less than the length of the correct-length run. Let A be the length of the correct-length run and $B_1, ..., B_5$ be the length of each incorrect run. Now, let's say we condition on $A = a$. We then have:

$$P(A > B_i | A = a) = P(B_i < a) = \Phi\left(\frac{a - 20}{\sqrt{2}}\right)$$

Now, the important thing is that conditioned on $A = a$, whether or not each B_i is greater than A is independent, so now we can multiply the probabilities:

$$P(A > B_1, ..., B_5 | A = a) = \Phi\left(\frac{a - 20}{\sqrt{2}}\right)^5$$

Ok, so the last step is to marginalize out a:

$$P(A > B_1, ..., B_5) = \int_{a=-\infty}^{\infty} f_A(a) \Phi\left(\frac{a - 20}{\sqrt{2}}\right)^5 da$$

where $f_A(a)$ is the pdf of $A \sim N(30, 3)$.

Because this solution is so nasty and doesn't use p, I think they were looking for p^5 even though it's wrong.

(e) See solution to Section 5 Problem 3 Part c.

(f) See solution to Section 5 Problem 3 Part d.

3. For parts a, c, and d, see PS5 #8 solutions. For part b, use definition of Beta to get 0.008.

For part e:

We've observed 10 successes and 10 fails from drug 1 (0.5 success rate), 75 successes and 5 fails from drug 2 (0.94 success rate). Under null hypothesis, we have one distribution from which we've observed 85 successes and 15 fails (0.85 success rate). To do our p-test, repeatedly draw one sample of size 20 and one sample of size 80 from Bernoulli(0.85) and see how often the difference in success rates is at lest 0.44. (See Corgi problem from section 6 if you're confused).

4. Let $Z = max(X, Y)$. There are two nice ways to solve for the PDF of Z:

   1. (X or Y must equal z, the other must be no higher): $f_Z(z) = f_X(z)P(Y \leq z) + f_Y(z)P(X \leq z) = 1 * z + 1 * z = 2z$

   2. (Find CDF, take derivative): $F_Z(z) = P(X < z, Y < z) = P(X < z)P(Y < z) = z^2 \implies f_Z(z) = \frac{dF_Z(z)}{dz} = 2z$

   We can then use our expectation formula:

$$E[Z] = \int_{z=0}^{1} z f_Z(z) = \int_{z=0}^{1} 2z^2 = \left[\frac{2z^3}{3}\right]_0^1 = \frac{2}{3}$$

5. We use "$\approx$" to denote where we make our Naive Bayes assumption:

$$\begin{aligned}\underset{y}{argmax} P(Y = y|X = x_1, x_2) &= \underset{y}{argmax} \frac{f(X_1 = x_1, X_2 = x_2|Y = y)P(Y = y)}{f(X_1 = x_1, X_2 = x_2)} \\ &\approx \underset{y}{argmax} \frac{f(X_1 = x_1|Y = y)f(X_2 = x_2|Y = y)P(Y = y)}{f(X_1 = x_1, X_2 = x_2)} \\ &= \underset{y}{argmax} f(X_1 = x_1|Y = y)f(X_2 = x_2|Y = y)P(Y = y)\end{aligned}$$

   Now note that we predict $Y = 1$ if:

$$f(X_1 = 5|Y = 1)f(X_2 = 3|Y = 1)P(Y = 1) > f(X_1 = 5|Y = 0)f(X_2 = 3|Y = 0)P(Y = 0)$$

   If I weren't so lazy, I would plug in the values from the problem at this point.

6. abc: See Section 8 Problem 3.

   For d:

Let $z_1 = \theta_1^{(y)} h_1$, $z_2 = \theta_2^{(y)} h_2$, and $z = z_1 + z_2$ so that $\hat{y} = \sigma(z)$:

$$
\begin{aligned}
\left| \frac{dLL}{dx_i} \right| &= \left| \frac{dLL}{\hat{y}} \frac{d\hat{y}}{dz} \frac{dz}{dx_i} \right| \\
&= \left| \frac{dLL}{\hat{y}} \frac{d\hat{y}}{dz} \left(\frac{dz_1}{dh_1} \frac{dh_1}{dx_i} + \frac{dz_2}{dh_2} \frac{dh_2}{dx_i} \right) \right| \\
&= \left| \left(\frac{y}{\hat{y}} - \frac{1-y}{1-\hat{y}} \right) \hat{y}(1-\hat{y}) \left(\theta_1^{(y)} (h_1(1-h_1)\theta_{i,1}^{(h)}) + \theta_2^{(y)} (h_2(1-h_2)\theta_{i,2}^{(h)}) \right) \right|
\end{aligned}
$$

e:

Divide θ_2 by 100. This makes the outputs of the dot products the same as before.