Final Spring 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. Every $\binom{P}{R}$ combination of people is equally likely.

   The number of combinations where you know exactly k people is $Q(k) = \binom{F}{k}\binom{P-F}{R-k}$ (as long as $k \leq F$). Note that $Q(k)$ is an integer-valued function, not a probability.

   Probability you know at least 6 people (using sample/event spaces) is:

$$1 - \frac{\Sigma_{k=0}^{5} Q(k)}{\binom{P}{R}} = 1 - \frac{\Sigma_{k=0}^{5} \binom{F}{k}\binom{P-F}{R-k}}{\binom{P}{R}}$$

2. (setup): Let J be getting the job, G_1 and G_2 be letters 1 and 2 being good. We then know:

   - $P(J|G_1G_2) = 0.75$
   - $P(J|G_1XORG_2) = 0.2$
   - $P(J|G_1^CG_2^C) = 0.05$
   - $P(G_1) = 0.8$
   - $P(G_2) = 0.5$

   NOTE: I'm just going to get each problem to the point where you can plug in from previous parts. Too lazy to plug in for you.

   (a) $P(G_1G_2) = P(G_1)P(G_2)$

   (b) $P(G_1XORG_2) = P(G_1)(1 - P(G_2)) + (1 - P(G_1))P(G_2)$

   (c) $P(G_1^CG_2^C) = (1 - P(G_1))(1 - P(G_2))$

$$P(J) = P(J|G_1G_2)P(G_1G_2) + P(J|G_1XORG_2)P(G_1XORG_2) + P(J|G_1^CG_2^C)P(G_1^CG_2^C)$$

(d) $P(G_1G_2|J) = \frac{P(J|G_1G_2)P(G_1G_2)}{P(J)}$

3. (a) Let X be #occurrences in 400 hours: $X \sim Poi(2)$, so $P(X < 5) = \Sigma_{k=0}^{4} P(X = k) = \Sigma_{k=0}^{4} \frac{2^k e^{-2}}{k!}$

   (b) See solution to PS3 Problem 9a.

   (c) See solution to PS3 Problem 9b.

4. (a) $\frac{1}{1+9^{\frac{1}{2}}} = \frac{1}{4}$

   (b) Note that $-E_B \sim N(1600, \frac{200^2}{2})$:

$$D = E_A - E_B = E_A + (-E_B) = N(1600, \frac{200^2}{2}) + N(-1600, \frac{200^2}{2}) = N(0, 200^2)$$

   (c)

$$\begin{aligned} P(\text{A wins}) &= \int_{x=-\infty}^{\infty} f_{E_A - E_B}(x) P(\text{A wins}|E_A - E_B = x) \\ &= \int_{x=-\infty}^{\infty} \frac{1}{\sqrt{2\pi 200^2}} e^{-\frac{x_i^2}{2*200^2}} \frac{1}{1 + 9^{-x}} \end{aligned}$$

5. (a) Using independence:

$$P(X < x, Y < y) = P(X < x)P(Y < y) = F_X(x)F_Y(y) = \Phi(x/2)\Phi(y/2)$$

   (b)

$$\begin{aligned} P(-0.5 < X < 0.5, -0.5 < Y < 0.5) &= P(-0.5 < X < 0.5)P(-0.5 < Y < 0.5) \\ &= (\Phi(0.5/2) - \Phi(-0.5/2))^2 \\ &= (\Phi(0.5/2) - (1 - \Phi(0.5/2))^2 \\ &= (2\Phi(0.5/2) - 1)^2 \end{aligned}$$

   and plug in using the Φ table.

6. (a) Code version of: $\sqrt{\frac{1}{n-1}\Sigma_{i=1}^{n}(S_i - \bar{S})^2}$ where $\bar{S} = \frac{\Sigma_{i=1}^{n} S_i}{n}$

   (b) Null hypothesis is that all data is drawn from the same distribution. To simulate the null hypothesis, we sample with replacement from all 200 grades. The pValue is the probability that, conditioned on the null hypothesis, the difference between the sample standard deviations would be more than 5 percentage points apart. To do find this, repeatedly draw two samples of size 100 (as described above) and see how often their sample variances are 5 percentage points or more apart. (See Corgi problem from section 6 if you're confused).

7. (a) Let $Y_i \sim Bernoulli(q)$:

$$E[Y_i] = P(Y_i = 1) = 0.5 * 0.5 + 0.5 * p_x = 0.25 + 0.5p_x = q$$

(b) $Var(Y_i) = q(1-q)$ (plug in from above)

(c) $\bar{Y} = \frac{\Sigma_{i=1}^{100} Y_i}{100} \sim \frac{Binom(100,q)}{100}$

(d) Note that (by our answer to part a), $p_x = 2q - 0.5$.

Since $\bar{Y}$ is an unbiased estimate of q (the sample mean is always an unbiased estimate), we have $p_x \approx 2\bar{Y} - 0.5$. It then makes sense to estimate $\hat{p_x} = 2\bar{Y} - 0.5$.

And this turns out to be unbiased:

$$E[\hat{p_x}] = 2E[\bar{Y}] - 0.5 = 2\frac{\Sigma_{i=1}^{100} E[Y_i]}{100} - 0.5 = 2q - 0.5 = p_x$$

(e) Assume for simplicity that $100q$ is an integer. I think my solution still works without this assumption, but it makes things easier. Then note that if $\bar{Y}$ is off of q by more than 0.05, $\hat{p_x}$ will be off by more than 0.1 (SEE NOTE), meaning that we need the sum of the Y_i's to be more than 5 away from $100q$ (using the fact that $\Sigma_{i=1}^{100} Y_i \sim Binom(100, q)$):

$$\begin{aligned} P(|\Sigma_{i=1}^{100} Y_i - 100q| > 5) &= 1 - P(100q - 5 \leq |\Sigma_{i=1}^{100} Y_i - 100q| \leq 100q + 5) \\ &= \Sigma_{i=100q-5}^{100q+5} P(Binom(100, q) = i) \end{aligned}$$

And then plug into the binomial PMF.
This solution is super janky. Maybe there's a cleaner one.
NOTE:

$$\begin{aligned} P(|\hat{p_x} - p_x| > 0.1) &= P(|2\bar{Y} - 0.5 - p_x| > 0.1) \\ &= P(|2\bar{Y} - 0.5 - 2q + 0.5| > 0.1) \\ &= P(|2\bar{Y} - 2q| > 0.1) = P(|\bar{Y} - q| > 0.05) \end{aligned}$$

8. (a)

$$\begin{aligned} \underset{\theta}{argmax} LL(\theta) &= \underset{\theta}{argmax} \Sigma_{i=1}^{N} \log(f_X(w_i)) \\ &= \underset{\theta}{argmax} \Sigma_{i=1}^{N} (\log(x) - \log(theta) - x^2/2\theta) \\ &= \underset{\theta}{argmax} (-N\log(\theta) - \frac{z}{2\theta}) \end{aligned}$$

where $z = \Sigma_{i=1}^{N} x^2$

We then take the derivative and set it to 0 to get our estimator:

$$\frac{dLL}{d\theta} = \frac{-N}{\theta} + \frac{z}{2\theta^2} = 0 \implies \theta = \frac{z}{2N} = \frac{\Sigma_{i=1}^{N} x^2}{2N}$$

The second derivative is

$$\frac{N}{\theta^2} - \frac{z}{\theta^3}$$

which is hopefully negative for our choice of θ, which would mean that we have found the maximum.

9: Section 8 Problem 2

10: Section 8 Problem 1