Chris Piech
CS 109

Problem Set #4
Oct 26, 2018

Problem Set #4 Solutions

With problems by Mehran Sahami and Chris Piech

1. a. Let X = the number of users that sign up for the social networking site in the next minute. We know that $X \sim \text{Poi}(5.5)$.

$P(X > 7) = 1 - P(X \leq 7) = 1 - \sum_{i=0}^{7} P(X = i) = 1 - \sum_{i=0}^{7} \frac{e^{-5.5} 5.5^i}{i!} \approx 0.1905$

b. Let Y = the number of users that sign-up for the social networking site in the next 2 minutes. We know that $Y \sim \text{Poi}(11)$.

$P(Y > 13) = 1 - P(Y \leq 13) = 1 - \sum_{i=0}^{13} P(Y = i) = 1 - \sum_{i=0}^{13} \frac{e^{-11} 11^i}{i!} \approx 0.2187$

c. Let W = the number of users that sign-up for the social networking site in the next 3 minutes. We know that $W \sim \text{Poi}(16.5)$.

$P(W > 15) = 1 - P(W \leq 15) = 1 - \sum_{i=0}^{15} P(W = i) = 1 - \sum_{i=0}^{15} \frac{e^{-16.5} 16.5^i}{i!} \approx 0.5820$

2. a. $1 = \int_{x=0}^{1} \int_{y=0}^{x} c\frac{y}{x} dy dx = c \int_{x=0}^{1} \frac{y^2}{2x}\Big|_0^x dx = c \int_{x=0}^{1} \frac{x^2}{2x} dx = c \int_{x=0}^{1} \frac{x}{2} dx = c\frac{x^2}{4}\Big|_0^1 = \frac{1}{4}c \Rightarrow$ $c = 4$

b. No, we cannot capture the constraint that $y < x$ in a factorization of the form $h(x)g(y)$.

c. $f_X(x) = \int_{y=0}^{x} \frac{4y}{x} dy = \frac{2y^2}{x}\Big|_0^x = \frac{2x^2}{x} = 2x$ where $0 < x < 1$

d. $f_Y(y) = \int_{x=y}^{1} \frac{4y}{x} dx = 4y \ln(x)\Big|_y^1 = -4y \ln(y)$ where $0 < y < 1$

e. $E[X] = \int_{x=0}^{1} x f_X(x) dx = \int_{x=0}^{1} x(2x) = \int_{x=0}^{1} 2x^2 = \frac{2x^3}{3}\Big|_0^1 = \frac{2}{3}$

f. $E[Y] = \int_{y=0}^{1} y f_Y(y) dy = \int_{y=0}^{1} y(-4y \ln(y)) = \int_{y=0}^{1} -4y^2 \ln(y) = -4 \int_{y=0}^{1} y^2 \ln(y) =$ $-4\Big[\frac{y^3}{3} \ln(y)\Big]\Big|_0^1 + 4 \int_{y=0}^{1} \frac{y^3}{3} \frac{1}{y} dy = 0 + 4 \int_{y=0}^{1} \frac{y^2}{3} dy = 4\Big[\frac{y^3}{9}\Big]\Big|_0^1 = \frac{4}{9}$

Note that we used integration by parts to evaluate the integral $\int_{y=0}^{1} y^2 \ln(y)$, where we have: $u = \ln(y), du = \frac{1}{y}dy, v = \frac{y^3}{3}$, and $dv = y^2 dy$.

3. a. For $X \sim \text{Uni}(a, b)$ we have $m = 0.5(b + a)$, computed as follows:

$$0.5 = F(m) = \int_a^m f(x) dx = \int_a^m \frac{1}{b-a} dx = \frac{x}{b-a}\Big|_a^m = \frac{m-a}{b-a}$$
$$m - a = 0.5(b - a)$$
$$m = 0.5(b + a)$$

b. For $X \sim N(\mu, \sigma^2)$, we have $m = \mu$, which readily follows from the fact that the Normal distribution is symmetric about the mean μ, so $F(\mu) = 0.5$

c. For $X \sim \text{Exp}(\lambda)$ we have $m = -\frac{\ln(0.5)}{\lambda}$, computed as follows:

$$\begin{aligned} 0.5 = F(m) &= 1 - e^{-\lambda m} \\ e^{-\lambda m} &= 0.5 \\ -\lambda m &= \ln(0.5) \\ m &= -\frac{\ln(0.5)}{\lambda} \end{aligned}$$

4. This is review from joint distributions.

a. Let T be you prior belief. The problem states that $T \sim N(98, 16)$. Thus:

$$f(t) = \frac{1}{4\sqrt{2\pi}} e^{\frac{-(t-98)^2}{2 \cdot 16}}$$

b. Let T be the true distance to the satellite. Let O be the observation of 100 a.u. The problem notes that $(O|T = t) = t + S$ where $S \sim N(0, 4)$. By linearity transform of a normal: $(O|T = t) \sim N(t, 4)$.

$$f(O = 100|T = t) = \frac{1}{2\sqrt{2\pi}} e^{\frac{-(100-t)^2}{2 \cdot 4}}$$

c. We can figure out our new PDF using Bayes Theorem of continuous variables

$$\begin{aligned} f(T = t|O = 100) &= \frac{f(O = 100|T = t) \cdot f(T = t)}{f(O = 100)} \\ &= K_1 \cdot \left(\frac{1}{2\sqrt{2\pi}} e^{\frac{-(100-t)^2}{2 \cdot 4}} \right) \cdot \left(\frac{1}{4\sqrt{2\pi}} e^{\frac{-(t-98)^2}{2 \cdot 16}} \right) \\ &= K_2 \cdot \left(e^{\frac{-(100-t)^2}{2 \cdot 4} + \frac{-(t-98)^2}{2 \cdot 16}} \right) \end{aligned}$$

Where K_2 is a constant such that the new belief integrates to 1. It is fine, but not necessary, to simplify the PDF further. Notice that in solving this we treat the denominator $f(O = 100)$ as a normalization constant and abstract it away as K_1.

5. As noted in class (or derived from the fact that $E[aX + b] = aE[X] + b$ and $\text{Var}(aX + b) = a^2\text{Var}(X)$), it follows that if $V \sim N(\mu, \sigma^2)$ then $(aV + b) \sim N(a\mu + b, a^2\sigma^2)$.

a. By convolution of two normals $A \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

b. By linearity of normal, $B \sim N(5\mu_1 + 2, 25\sigma_1^2)$

c. Applying the linearity of normal property to $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$, and $Z \sim N(\mu_3, \sigma_3^2)$, we have respectively $(aX) \sim N(a\mu_1, a^2\sigma_1^2), (-bY) \sim N(-b\mu_2, b^2\sigma_2^2), (c^2Z) \sim N(c^2\mu_3, c^4\sigma_3^2)$. Then noting that $C = aX - bY + c^2Z = aX + (-bY) + c^2Z$ and using the fact that the sum of Normal distributions is a Normal distribution whose parameters are the sum of the respective parameters of the summed distributions, we obtain the answer: $C \sim N(a\mu_1 - b\mu_2 + c^2\mu_3, a^2\sigma_1^2 + b^2\sigma_2^2 + c^4\sigma_3^2)$

6. a. Let $D = A_1 - A_2$ be the difference between the two random variables. It is the convolution of A_1 and $(-A_2)$. Thus $D \sim N(\mu = S_1 - S_2, \sigma^2 = 2(\frac{2000}{7})^2)$.
   b. $P(D > 0) = 1 - P(D < 0) = 1 - \Phi(\frac{0-(S_1-S_2)}{404.06}) = \Phi(\frac{S_1-S_2}{404.06})$
   c. The probability that Ke wins is:
   $p = 1 - \Phi(\frac{S_2-S_1}{404.06}) = 1 - \Phi(3.7865) = 0.0000764$.
   The number of wins is a binomial $X \sim \text{Bin}(n, 0.0000764)$. We want $E[X] > 1$ which implies that $0.0000764n > 1$. For this inequality to hold, $n > 13089$.

7. (a) $P(X = j, Y = i) = \frac{1}{6j}$ where $j = 1, 2, ..., 6$ and $i = 1, 2, ..., j$. We can choose any of the 6 values for j (with equal probability $\frac{1}{6}$) and then for i we can choose any of the j values in the subset $1, 2, ..., j$ (with equal probability $\frac{1}{j}$).
   (b) $P(X = j|Y = i) = \dfrac{P(X = j, Y = i)}{P(Y = i)} = \dfrac{\frac{1}{6j}}{\sum_{k=i}^{6} \frac{1}{6k}} = \dfrac{\frac{1}{j}}{\sum_{k=i}^{6} \frac{1}{k}}$ where $j = 1, 2, ..., 6$ and $i = 1, 2, ..., j$. Note that $P(Y = i) = \displaystyle\sum_{k=i}^{6} P(X = k, Y = i) = \sum_{k=i}^{6} \frac{1}{6k}$ where the bounds on the sum are determined by the fact that the value of X (indexed by k) must be greater than or equal to the value of Y (denoted by i).
   (c) No, X and Y are not independent (they are dependent), as the possible values for Y depend on the initially chosen value for X. For example if $X = 1$, then we know $Y = 1$.

8. Note that $\sigma^2 = 52900$ yields $\sigma = 230$.
   a. Let $W = X_1 + X_2$ = number of visitors to website in two weeks. Since W is a sum of Normal distributions, we have: $W \sim N(2(2200), 2(52900)) = N(4400, 105800)$. Thus the standard deviation of W, $\sigma_W = \sqrt{105800} = 325.27$. Now, we compute $P(W > 5000)$:

$$P(W > 5000) = 1 - P(W \leq 5000) = 1 - P(W < 5000.5) = 1 - P\left(\frac{W - 4400}{325.27} \leq \frac{5000.5 - 4400}{325.27}\right)$$
$$= 1 - P(Z < 1.85) = 1 - 0.9678 \approx 0.0322$$

   Note that we must do a continuity correction because we are using a continuous distribution to represent a discrete quantity (number of users).
   b. First, compute:
   $P(X > 2000) = 1 - P(X \leq 2000) = 1 - P(X < 2000.5) = 1 - P\left(\frac{X-2200}{230} \leq \frac{2000.5-2200}{230}\right)$
   $= 1 - P(Z \leq -0.87) = P(Z \leq 0.87) \approx 0.8078$. So, let $p = 0.8078$.
   Let Y = number of weeks that $X > 2000$ in the next three weeks. Since the number of visitors to the web site each week is independent, we have $Y \sim Bin(3, p)$ and we need to compute $P(Y = 2) + P(Y = 3) = \binom{3}{2}p^2(1 - p) + \binom{3}{3}p^3 \approx 0.9034$

9. The joint density of the point (X, Y) where the package is dropped off is: $f_{X,Y}(x, y) = \frac{1}{100} = \frac{1}{10} \cdot \frac{1}{10} = f_X(x)f_Y(y)$, where $-5 < x, y < 5$, since the point is uniformly distributed in robot's square world. Noting that the joint density function factors as separate functions of X and Y, we know that X and Y are independent and $X \sim Y \sim \text{Uni}(-5, 5)$. Thus:

$$E[D] = E[|X| + |Y|] = \int_{x=-5}^{5} |x|\frac{1}{10}dx + \int_{y=-5}^{5} |y|\frac{1}{10}dy = 2\int_{x=0}^{5} x\frac{1}{10}dx + 2\int_{y=0}^{5} y\frac{1}{10}dy$$
$$= 2\frac{x^2}{20}\Big|_{x=0}^{5} + 2\frac{y^2}{20}\Big|_{y=0}^{5} = 2\frac{25}{20} + 2\frac{25}{20} = 5$$

10. Let event A_i = "max update" occurs on iteration i of the **for** loop. Let indicator variable $X_i = 1$ if the event A_i occurs, 0 otherwise. Now note that the element at position i has a $\frac{1}{i}$ probability of having a greater value than all values before it in the array. This follows from the fact that we can think of the maximal value of the first i array elements being equally likely to be in any of the first i array positions. Thus, $\mathrm{P}(A_i) = \frac{1}{i}$. Now compute:
$$E\Big[\sum_{i=1}^{n} X_i\Big] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{n} \frac{1}{i}$$
(As a side note, $\sum_{i=1}^{n} \frac{1}{i} = O(\log n)$. So while the overall running time of the **max** function is clearly $O(n)$, the number of times that the variable **max** is actually updated is only $O(\log n)$.)

11. a. Note that for $X \sim \mathrm{Beta}(a, b)$, we have:
    $E[X] = 0.5 = \frac{a}{a+b} \Rightarrow a = b$ and $\mathrm{Var}(X) = \frac{1}{36} = \frac{ab}{(a+b)^2(a+b+1)}$
    This gives us two equations and two unknowns (a and b). Substituting $a = b$ into the equation for $\mathrm{Var}(X)$, we have $\frac{ab}{(a+b)^2(a+b+1)} = \frac{a^2}{(2a)^2(2a+1)} = \frac{1}{8a+4} = \frac{1}{36}$. Solving for a we obtain $a = 4$ and then set $b = a = 4$.
    b. As derived in class, the Beta distribution is conjugate to itself so if our prior (starting) distribution is a Beta, then so will be our posterior distribution (after observing the coin flips). Our prior distribution is
    Beta(4, 4), so after observing 8 heads and 5 tails we have a posterior distribution that is Beta(12, 9).
    c. We know that the parameters of our posterior Beta distribution (after observing the 8 heads and 4 tails) are $a = 12$ and $b = 8$.
    So $E[X|12$ flips resulting in 8 heads and 4 tails$] = \frac{12}{12+8} = \frac{12}{20} = \frac{3}{5}$.
    d. We know that the parameters of our posterior Beta distribution (after observing the 8 heads and 4 tails) are $a = 12$ and $b = 8$.
    So $\mathrm{Var}(X|12$ flips resulting in 8 heads and 4 tails$) = \frac{ab}{(a+b)^2(a+b+1)} = \frac{(12)(8)}{(12+8)^2(12+8+1)} \approx 0.0114$

12. Here are the Titanic probabilities and statistics:
    a. $P(S = \text{true}|G = \text{female}, C = 1) = 0.97$
    $P(S = \text{true}|G = \text{female}, C = 2) = 0.92$
    $P(S = \text{true}|G = \text{female}, C = 3) = 0.50$
    $P(S = \text{true}|G = \text{female}, C = 1) = 0.37$
    $P(S = \text{true}|G = \text{female}, C = 2) = 0.16$
    $P(S = \text{true}|G = \text{female}, C = 3) = 0.14$
    b. You could have used any Beta prior when calculating your answer (as long as you were clear which one you used). We assumed a uniform prior. 22 survived, 31 perished.
    $p \sim \mathrm{Beta}(23, 32)$
    c. The expected fares for the different classes are:
    Class 1: £84.15

Class 2: £20.66
Class 3: £13.71

13. a. $E[X] = 7.4$, $E[Y] = 8.0$
b. $E[X^2] = 58.8$, $E[Y^2] = 68.2$
c. $X \sim N(7.4, 3.98)$, $Y \sim N(8.0, 3, 68)$
d. Person B is 32 times more likely to have written the email than Person A. Let A be the event that person A wrote the email. Let B be the event that person B wrote the email. Let O be the event that we observed the given keystroke timings in the email.

$$
\begin{aligned}
\frac{P(A|O)}{P(B|O)} &= \frac{\frac{P(O|A)P(A)}{P(O)}}{\frac{P(O|B)P(B)}{P(O)}} && (1)\\
&= \frac{P(O|A)}{P(O|B)} && (2)\\
&= \frac{\prod_i P(X_i|A)}{\prod_i P(X_i|B)} && (3)\\
&= \frac{\prod_i f(X_i|A) \cdot \epsilon}{\prod_i f(X_i|B) \cdot \epsilon} && (4)\\
&= \frac{\prod_i f(X_i|A)}{\prod_i f(X_i|B)} && (5)\\
&= \frac{\prod_i \frac{1}{\sigma_A} e^{\frac{-(x-\mu_A)^2}{2\sigma_A^2}}}{\prod_i \frac{1}{\sigma_B} e^{\frac{-(x-\mu_B)^2}{2\sigma_B^2}}} && (6)\\
&\approx 0.031 && (7)
\end{aligned}
$$

Thus,
$\dfrac{P(B|O)}{P(A|O)} \approx \dfrac{1}{0.031} \approx 32$

Explanations:
(1) Bayes theorem
(2) Since P(A) = P(B)
(3) Since each keystroke timing is independent
(4) This approximates the probability
(5) Epsilons cancel
(6) PDF of normals
(7) Using Python