STAT 391 Midterm Solutions
Thursday May 8 2014, 11:30-12:20

- 3 pages of notes allowed
- no other sources of information are allowed
- electronic devices are not allowed

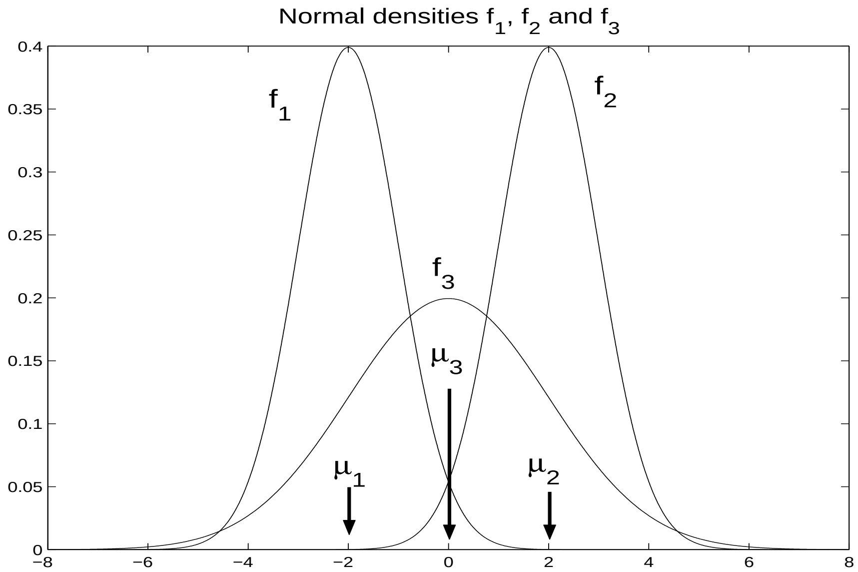
- *Do Well!*

Problem 1	4 points	(Normal distributions)
Problem 2	5 points	(Discrete probabilities)
Problem 3	5.5 points	(Properties of Mean and Variance)
Problem 4	4.5 points	(ML estimation with missing info)
Bonus	1 point	
Total	20 points	
Extra credit	3 points	

4 points

Problem 1 – Normal distribution

The graph below shows three normal densities f_1, f_2, f_3 having parameters (μ_1, σ_1), (μ_2, σ_2), (μ_3, σ_3), and denote by P_1, P_2 and P_3 the probability distributions associated with f_1, f_2 and f_3, respectively.

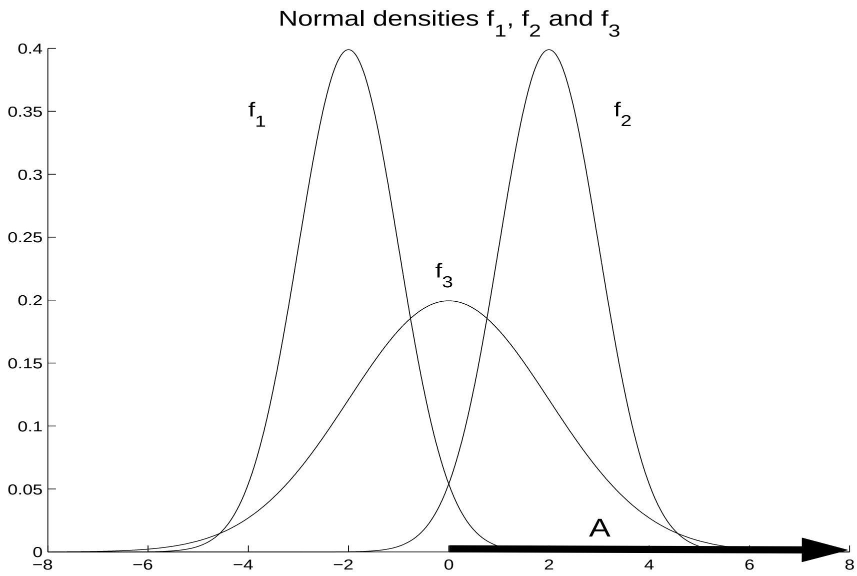


a. Which density has the largest mean μ ? A: f_2.

b. Which density has the largest standard deviation σ ? A: f_3

c. Mark on the graph shown above the positions of μ_1, μ_2, μ_3. See graph.

d. Let A denote the event $x \geq \mu_3$. Draw A on the x axis of the figure below (this is the same figure as on the previous page).



e. $P_3(A)$ the probability of event A under the distribution represented by f_3 is (choose one):

☐ 0	☐ close to 0	☑ √ 1/2	☐ close to 1/2
☐ 1	☐ close to 1		

f. $P_1(A)$ the probability of event A under the distribution represented by f_1 is (choose one):

☐ 0	☑ √ close to 0	☐ 1/2	☐ close to 1/2
☐ 1	☐ close to 1		

g. $P_2(A)$ the probability of event A under the distribution represented by f_2 is (choose one):

☐ 0	☐ close to 0	☐ 1/2	☐ close to 1/2
☐ 1	☑ √ close to 1		

h. Is the statement "$P_1(A) = 1 - P_2(A)$" true or false? A: TRUE

5 points

Problem 2 – Discrete distributions *Show your work*

Pilgrims come from all over the world to consult the famous Oracle of

Randonesia. The Oracle answers each question with one of the numbers $\{-1, +1, 0\}$ (which are interpreted as "no", "yes", "maybe" according to the local priests). Chung is the youngest priest in training, and he has come to think that the Oracle answers each question randomly and independently from a distribution P defined by parameters $(\theta_-, \theta_0, \theta_+)$.

3 points

2.a. Let $a \in \{-1, +1, 0\}$ be the Oracle's answer to a question from a pilgrim. Assuming Chung is right, give answers to the following probability questions as functions of $(\theta_-, \theta_0, \theta_+)$.

$P(a < 1) = P(a \neq 1) = P(a = 0\, OR\, a = -1) = P(a = 0) + P(a = 1) = \theta_0 + \theta_-$

Three pilgrims receive answers a_1, a_2, a_3. What is the probability that these answers are all different?
A:There are $3! = 6$ different ways to get all different answers. Since we must have one answer of each, all three probabilities will be multiplied together. This gives us an answer of $3!\theta_0\theta_-\theta_+$

Another group of three pilgrims receive answers a_4, a_5, a_6 from the Oracle. What is the probability that $a_4 = a_5 = a_6 = 0$?
A: All three must get 0. There is only one way to do this. Thus, we mulitply the probabilities together to get θ_0^3

What is the probability that $a_4 = a_5 = a_6$?
A: This is similar to the previous question, except there are three ways for all three pilgrims to get the same answer: all 0's, all 1's, and all -1's. Adding these three probabilities together (since they are disjoint events) we get $\theta_0^3 + \theta_-^3 + \theta_+^3$.

1.5 points

2.b. Assume $\theta_0 = \frac{1}{2}$, $\theta_- = \frac{1}{4}$, $\theta_+ = \frac{1}{4}$ (and that Chung is right).
Calculate $E[a]$. A: $E[a] = 0 * \frac{1}{2} + 1 * \frac{1}{4} + (-1) * \frac{1}{4} = 0$

Calculate $E[a^2]$. A: $E[a^2] = 0^2 * \frac{1}{2} + 1^2 * \frac{1}{4} + (-1)^2 * \frac{1}{4} = \frac{1}{2}$

0.5 points

2.c. The following day of $n = 10$ pilgrims arrive, and receive the following answers. $\mathcal{D} = \{-1, -1, -1, 1, 1, 0, 0, 0, 0, 0, \}$. What are the Maximum Likelihood estimates of $(\theta_-, \theta_0, \theta_+)$ from this data set? *No need to show work. Numerical result sufficient.* A: $\theta_0^{ML} = \frac{5}{10}$, $\theta_+^{ML} = \frac{2}{10}$, $\theta_-^{ML} = \frac{3}{10}$

1 point

[**d. – Extra credit**] Chung's friend Kai Lai believes that the Oracle answers randomly and independently from another distribution $\tilde{P}$ on $\{-1, +1, 0\}$. $\tilde{P}$ has parameters $(\tilde{\theta}_-, \tilde{\theta}_0, \tilde{\theta}_+)$ with the property that $\tilde{\theta}_- = \tilde{\theta}_+ = \tilde{\theta}_1$. Write the likelihood of $(\tilde{\theta}_0, \tilde{\theta}_1)$ given the dataset $\mathcal{D}$ from the previous question.

A: The likelihood can be written as $\tilde{\theta}_0^5\tilde{\theta}_1^5$. However, we have the relationship that $\tilde{\theta}_0 + 2\tilde{\theta}_1 = 1$. So, we can rewrite the likelihood as $\tilde{\theta}_0^5\left(\frac{1-\tilde{\theta}_0}{2}\right)^5$.

1 point

[e. – Extra credit] Find now the expressions and numerical values of $(\tilde{\theta}_-, \tilde{\theta}_0, \tilde{\theta}_+)$ by maximizing the likelihood obtained in **d.**.
A: The log likelihood of the above is $5\log(\tilde{\theta}_0)+5\log(1-\tilde{\theta}_0)-5\log(2)$. Taking the derivative with respect to $\tilde{\theta}_0$, we get $\frac{5}{\tilde{\theta}_0} - \frac{5}{1-\tilde{\theta}_0} = 0$. This can be solved for $\tilde{\theta}_0^{ML} = \frac{1}{2}$. We then get that $\tilde{\theta}_1^{ML} = \tilde{\theta}_+ = \tilde{\theta}_- = \frac{1}{4}$.

1 point

[f. – Extra credit] *No need to show work.* Mark the correct answer.
■ $l(\tilde{\theta}_-, \tilde{\theta}_0, \tilde{\theta}_+) < l(\theta_-, \theta_0, \theta_+)$ for all datasets
☐ $l(\tilde{\theta}_-, \tilde{\theta}_0, \tilde{\theta}_+) = l(\theta_-, \theta_0, \theta_+)$ for all datasets
☐ $l(\tilde{\theta}_-, \tilde{\theta}_0, \tilde{\theta}_+) > l(\theta_-, \theta_0, \theta_+)$ for all datasets
■ None of the above.
In this question, technically the correct answer is "none of the above", though points were given for the answer <. The likelihood of the constrained model will be less than OR EQUAL TO the likelihood of the general multinomial. The reason is that the likelihoods are equal when the counts of 1's and −1's are the same. Otherwise, the constraint that $\tilde{\theta}_- = \tilde{\theta}_+$ means the likelihood of the data will be less than that of the unconstrained multinomial.

5.5 points

Problem 3 – Properties of Mean and Variance *Show your work for all but a*

X is a random variable with $E[X] = -1$, $Var(X) = 1$.

1.5 points

a. Y is another variable with $E[Y] = E[X]$, $Var(Y) = Var(X)$. Mark the correct answer to the questions below

$Y = X$	☐ True	■ False
$S_Y = S_X$	☐ True	■ False
Y is $Normal(-1, 1)$	☐ True	■ False

1 point

b. Let $Z = X + 1$. What is $E[Z]$? What is $Var(Z)$?
A: $E[Z] = E[X + 1] = E[X] + 1 = -1 + 1 = 0$.
$Var(Z) = Var(X + 1) = Var(X) = 1$.

1 point

c. Let $U = (X + 1)^2$. What is $E[U]$?
A: $E[U] = E[(X + 1)^2] = E[(X - (-1))^2] = E[(X - E[X])^2] = Var(X) = 1$

2 points

d. $W = X^2$. What is $E[W]$?
A: $E[W] = E[X^2] = Var(X) + E[X]^2 = 1 + (-1)^2 = 2$.

4.5 points

Problem 4 – ML estimation with missing information *Show your work*

You record n samples from a geometric distribution with parameter γ. However, due to a mistake, all that ends up being recorded is whether each sample was zero or non-zero.

The Geometric distribution on $S = \{0, 1, 2, \ldots n, \ldots\}$ has density $f_\gamma(n) = (1-\gamma)\gamma^n$, $\gamma \in (0, 1)$

.5 points

a. What is the outcome space S' of this experiment? (Find an appropriate symbol for "non-zero".) What is the probability of each outcome in S'?
A: $S' = \{0, 1\}$ where 0 is the outcome of sampling a zero and 1 is the outcome of sampling a non-zero value. Since the probability of getting 0 is the same in either model space, we have $P(0) = 1 - \gamma$. In our new space, the only outcomes are 1 and 0, thus $P(1) = 1 - P(0) = 1 - (1 - \gamma) = \gamma$.

2 points

b. Write the log-likelihood of the data as a function of γ, using the probabilities you obtained in **a.**. What are the sufficient statistics?
A: Let X_i be the i^{th} outcome. The likelihood is $\prod_{i=1}^{n}(1-\gamma)^{1-X_i}\gamma^{X_i} = (1-\gamma)^{c_0}\gamma^{c_1}$ where c_0 and c_1 are the counts of 0's and 1's respectively. Taking logs the log-likelihood is $c_0 \log(1-\gamma) + c_1 \log(\gamma)$. Here, c_0 and c_1 are the sufficient statistics in this problem because the data is reduced to them when we compute likelihoods. Alternatively, the likelihoods of any sequences with the same number of counts are indistinguisable from each other.

2 points

c. Now find the expression of maximum of the log-likelihood in **b.** and thus derive a formula for Maximum Likelihood estimate of γ.
A: Taking the derivative of the log-likelihood gives us $-\frac{c_0}{1-\gamma} + \frac{c_1}{\gamma}$. Setting this equal to 0 and solving gives us $\gamma^{ML} = \frac{c_1}{c_0+c_1} = \frac{c_1}{n}$ where n is the total number of trials.