....

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

Student name: ....

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

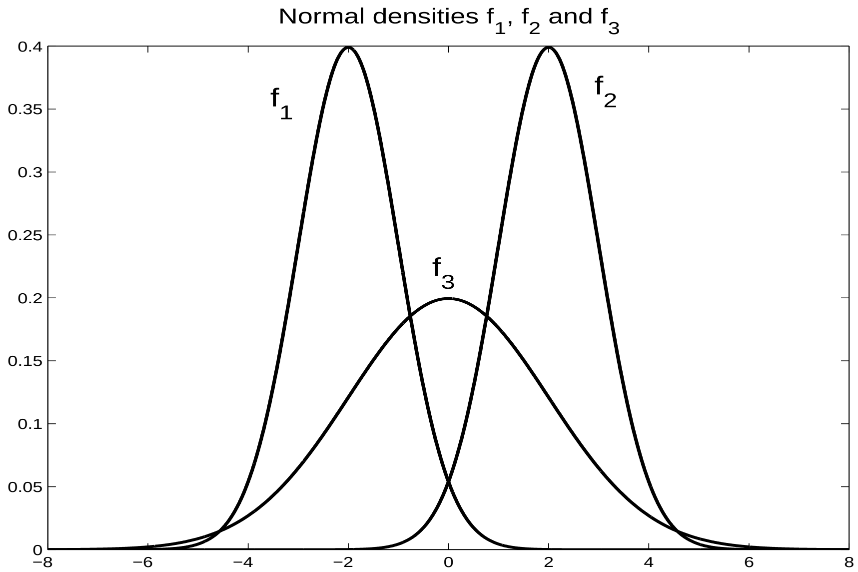
- *Do Well!*

Problem 1	4 points	(Normal distributions)
Problem 2	2 points	(Discrete probabilities)
Problem 3	4 points	(Properties of Expectation and Variance)
Problem 4	8.5 points	(ML estimation with missing information)
Bonus	0.5 points	
Total	24 points	
Extra credit	3 points	

5 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.

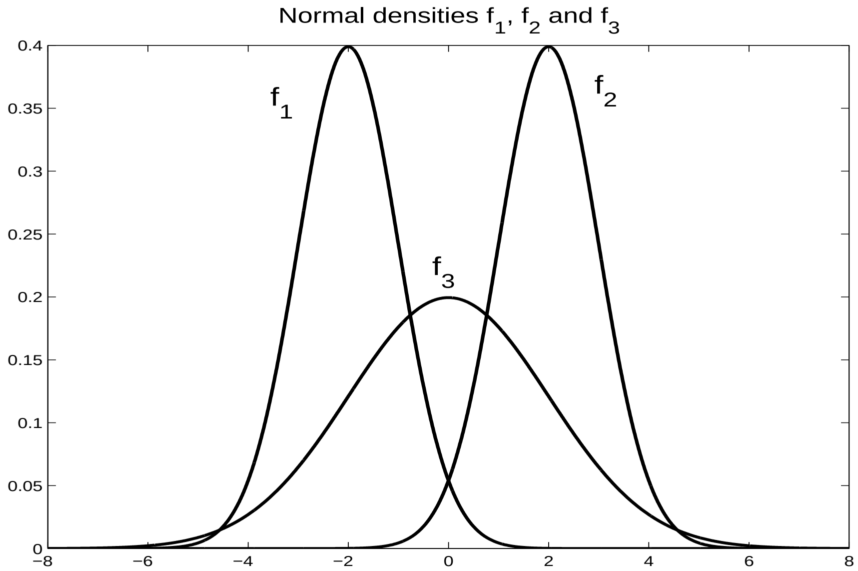


a. Which density has the largest mean μ ?

b. Which density has the largest standard deviation σ ?

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

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?

Problem 2 – Discrete distributions<u>*Show your work*</u>

Let $S = \{-1, 0, 1\}$ be the outcome space of an experiment, and P defined by $(\theta_-, \theta_0, \theta_+)$ a distribution over S.

a. Give answers to the following questions as functions of $(\theta_-, \theta_0, \theta_+)$.

$P(X < 1) =$

(X_1, X_2, X_3) are three independent samples from P. What is the probability that they are all different?

(X_1, X_2, X_3) are three independent samples from P. What is the probability that $X_1 = X_2 = X_3 = 0$?

(X_1, X_2, X_3) are three independent samples from P. What is the probability that $X_1 = X_2 = X_3$?

b. Assume $\theta_0 = \frac{1}{2}$, $\theta_- = \frac{1}{4}$, $\theta_+ = \frac{1}{4}$.
Calculate $E[X]$.

Calculate $E[X^2]$.

c. The following dataset of $n = 10$ points was sampled i.i.d. from P. $\mathcal{D} = \{-1, -1, -1, 1, 1, 0, 0, 0, 0, 0, \}$. What are the Maximum Likelihood estimates of $(\theta_-, \theta_0, \theta_+)$ from this data set? *No need to show work. Numerical result sufficient.*

[**d. – Extra credit**] Assume now that instead of P, there is another distribution $\tilde{P}$ on S from which $\mathcal{D}$ was sampled. $\tilde{P}$ has parameters $(\tilde{\theta}_-,\, \tilde{\theta}_0,\, \tilde{\theta}_+)$ with the property that $\tilde{\theta}_- = \tilde{\theta}_+ = \tilde{\theta}_1$. (In other words, the parameter of the model $\tilde{P}$ are *tied*, or *have constraints*.) Write the likelihood of $(\tilde{\theta}_0, \tilde{\theta}_1)$ given the dataset $\mathcal{D}$.

[**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.**.

[**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.

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

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

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

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

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

d. $W = X^2$. What is $E[W]$?

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

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?

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?

c. Now find the expression of maximum of the log-likelihood in **b.** and thus derive a formula for Maximum Likelihood estimate of γ.