Midterm: ARE 210

Fall 2015

Due: Wednesday, October 28, 2015

This exam is due by 5pm on Wednesday, October 28 at the start of lecture. No late exams will be accepted.

You may use *Introduction to Mathematical Statisitics* by Hogg, Craig, and McKean, DeGroot and Schervish's *Probability and Statistics*, Alder's *R in a Nutshell*, my lecture slides, this semester's problem sets and solutions, your own notes, and `R`'s built-in help system. You may not use any other resources, including notes or other materials from previous editions of this course. You may not discuss this exam with anyone until after the due date. Any violations of these policies will result in a 0 for this exam and could result in failure of the course and further disciplinary action.

If questions are unclear, you can e-mail me. Any clarification(s) that I may offer will be distributed via the class-wide e-mail list.

You must show your work for full credit. Any results given or proven in lecture or on a problem set can be used directly; otherwise, you should prove it for full credit. If you are not able to prove a result, partial credit will be given for logical explanations.

1 (13 pts) Transformation relations

Let $X_i \sim \text{Uniform}[0, 1], i = 1, \ldots, N$. Let $Y_i = \exp(X_i)$.

1. (4 pts) Derive the correlation between X and Y.

   Solution: First, recognize that $f(x) = 1$ for $x \in [0, 1]$ and 0 otherwise. Then, we have:

$$
\begin{aligned}
\mathbb{E}[X] &= \int_0^1 x\,dx = \frac{1}{2} \\
\mathbb{E}\left[X^2\right] &= \int_0^1 x^2\,dx = \frac{1}{3} \\
\text{Var}(X) &= \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12} \\
\mathbb{E}[Y] &= \int_0^1 \exp(x)\,dx = e - 1 \\
\mathbb{E}[Y^2] &= \int_0^1 \exp(2x)\,dx = \frac{1}{2}\exp(2x)\Big|_{x=0}^{1} = \frac{\exp(2)-1}{2} \\
\text{Var}(Y) &= \frac{exp(2)-1}{2} - (e-1)^2 \\
\mathbb{E}[XY] &= \int_0^1 x\exp(x)\,dx = x\exp(x)|_{x=0}^{1} - \int_0^1 \exp(x)\,dx = \exp(1) - (\exp(1)-1) = 1 \\
\text{Cov}(X,Y) &= 1 - \frac{1}{2}(e-1) = \frac{3-e}{2} \\
\text{Cor}(X,Y) &= \frac{3-e}{2}\left[\frac{1}{12}\right]^{-1/2}\left[\frac{\exp(2)-1}{2} - (e-1)^2\right]^{-1/2}
\end{aligned}
$$

Note that the sixth line uses integration by parts.

Consider the vector (x_i, y_i) paired to (x_j, y_j). The pair is *concordant* if $x_i < x_j$ and $y_i < y_j$ or $x_i > x_j$ and $y_i > y_j$. The pair is *discordant* if $x_i < x_j$ and $y_i > y_j$ or $x_i > x_j$ and $y_i < y_j$. A pair is neither if $x_i = x_j$ or $y_i = y_j$.

2. (2 pts) How many unique pairs are possible?
   Solution: There are $\binom{N}{2}$ unique pairs.

Consider the ratio

$$\frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{\text{total number of pairs}}$$

3. (1 pt) What is the range of this ratio?
   Solution: The range is $[-1, 1]$.

4. (2 pts) What is the value of this ratio for X_i and Y_i?
   Solution: It is 1.

5. (2 pts) Under what circumstances does this ratio take on its minimum and maximum values?

Solution: It is 1 if one variable is a monotonically increasing function of the other; it is -1 if one variable is a monotonically decreasing function of the other. In other words, it takes on the extreme values if the ordering of one random variable is perfectly maintained or perfectly reversed in comparison to the other.

6. (2 pts) When does (the standard measure of) correlation take on its minimum and maximum values?

   Solution: It is 1 if one variable is a linear function with positive slope of the other; it is -1 if one variable is a linear function iwth negative slope of the other.

2 (20 pts) Contamination implications

Suppose that $X_1, \ldots, X_N$ are i.i.d. with mean μ and variance σ^2. Suppose that X_{N+1} is contaminated such that $X_{N+1} = \mu + k$, both μ and k unknown, but not random.

Define:

$$\bar{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$\bar{X}_{N+1} = \frac{1}{N+1}\sum_{i=1}^{N+1} X_i$$

$$\hat{\sigma}_N^2 = \frac{1}{N-1}\sum_{i=1}^{N} \left(X_i - \bar{X}_N\right)^2$$

$$\hat{\sigma}_{N+1}^2 = \frac{1}{N}\sum_{i=1}^{N+1} \left(X_i - \bar{X}_{N+1}\right)^2.$$

1. (4 pts) Write $\bar{X}_{N+1}$ as a function of $\bar{X}_N$.

   Solution: We have:

$$\bar{X}_{N+1} = \frac{N}{N+1}\bar{X}_N + \frac{1}{N+1}X_{N+1}.$$

2. (2 pts) Find the expected value of $\bar{X}_{N+1}$ as a function of μ.

   Solution: Using the result in the previous part, we have:

$$\mathbb{E}\left[\bar{X}_{N+1}\right] = \mu + \frac{k}{N+1}.$$

3. (4 pts) Write $\hat{\sigma}_{N+1}^2$ as a function of $\hat{\sigma}_N^2$.

Solution: We find:

$$
\begin{aligned}
\hat{\sigma}^2_{N+1} &= \frac{1}{N}\sum_{i=1}^{N+1}\left(X_i - \bar{X}_{N+1}\right)^2 \\
&= \frac{1}{N}\sum_{i=1}^{N+1}\left(X_i - \bar{X}_N + \bar{X}_N - \bar{X}_{N+1}\right)^2 \\
&= \frac{1}{N}\left[\sum_{i=1}^{N+1}\left(X_i - \bar{X}_N\right)^2 + \sum_{i=1}^{N+1}\left(\bar{X}_N - \bar{X}_{N+1}\right)^2 + 2\sum_{i=1}^{N+1}\left(X_i - \bar{X}_{N+1}\right)\left(\bar{X}_N - \bar{X}_{N+1}\right)\right] \\
&= \frac{1}{N}\Bigg[(N-1)\hat{\sigma}^2 + \left(X_{N+1} - \bar{X}_N\right)^2 + (N+1)\left(\bar{X}_N - \bar{X}_{N+1}\right)^2 \\
&\quad + 2\left(\bar{X}_N - \bar{X}_{N+1}\right)\sum_{i=1}^{N+1}\left(X_i - \bar{X}_{N+1}\right)\Bigg] \\
&= \frac{1}{N}\left[(N-1)\hat{\sigma}^2 + \left(X_{N+1} - \bar{X}_N\right)^2 + (N+1)\left[\frac{X_{N+1} - \bar{X}_N}{N+1}\right]^2 - \frac{2}{N+1}\left(X_{N+1} - \bar{X}_N\right)^2\right] \\
&= \frac{N-1}{N}\hat{\sigma}^2 + \frac{1}{N+1}\left(X_{N+1} - \bar{X}_N\right)^2 \\
&= \frac{N-1}{N}\hat{\sigma}^2 + \frac{1}{N+1}\left(\mu + k - \bar{X}_N\right)^2 \\
&= \frac{N-1}{N}\hat{\sigma}^2 + \frac{1}{N+1}\left[\left(\bar{X}_N - \mu\right)^2 - 2k\left(\bar{X}_N - \mu\right) + k^2\right]
\end{aligned}
$$

The last two lines will be helpful for the next part.

4. (2 pts) Find the expected value of $\hat{\sigma}^2_{N+1}$ as a function of σ^2.
Solution: We start with knowing that $\mathbb{E}\left[\hat{\sigma}^2\right] = \sigma^2$. Then:

$$
\begin{aligned}
\mathbb{E}\left[\hat{\sigma}^2_{N+1}\right] &= \mathbb{E}\left[\frac{N-1}{N}\hat{\sigma}^2 + \frac{1}{N+1}\left[\left(\bar{X}_N - \mu\right)^2 - 2k\left(\bar{X}_N - \mu\right) + k^2\right]\right] \\
&= \frac{N-1}{N}\sigma^2 + \frac{1}{N+1}\left[\mathbb{E}\left[\left(\bar{X}_N - \mu\right)^2\right] - 2k\mathbb{E}\left[\bar{X}_N - \mu\right] + k^2\right] \\
&= \frac{N-1}{N}\sigma^2 + \frac{1}{N+1}\left[\mathrm{Var}\left(\bar{X}_N\right) + k^2\right] \\
&= \frac{N-1}{N}\sigma^2 + \frac{\sigma^2}{N(N+1)} + \frac{k^2}{N+1} \\
&= \frac{(N+1)(N-1)+1}{N(N+1)}\sigma^2 + \frac{k^2}{N+1} \\
&= \frac{N^2}{N(N+1)}\sigma^2 + \frac{k^2}{N+1} \\
&= \frac{N}{N+1}\sigma^2 + \frac{k^2}{N+1}.
\end{aligned}
$$

In practice, the researcher does not know which observation is contaminated (*i.e.*, he doesn't know which observation is the $N+1$st) or even whether a value is contaminated. Instead, he uses an outlier detection procedure. Specifically, he classifies as an outlier any observation that is more than c standard deviations away from the mean; *i.e.*, observation i is an outlier if

$$\left|X_i - \bar{X}_{N+1}\right| > c\sqrt{\hat{\sigma}^2_{N+1}}.$$

5. (8 pts) Find the values of k that would result in an observation being rejected as a function of $\hat{\sigma}^2_N$ and c.

   Solution: Consider the expected values of these quantities. Then, the test rejects if the square of the difference exceeds the square of the threshold:

$$\left(\mu + k - \mu - \frac{k}{N+1}\right)^2 > c^2\left[\frac{N}{N+1}\sigma^2 + \frac{k^2}{N+1}\right]$$

$$\left(\frac{N^2}{(N+1)^2} - \frac{c^2}{N+1}\right)k^2 > c^2\frac{N}{N+1}\sigma^2$$

$$|k| > c\sigma\sqrt{\left[\left(\frac{N^2}{(N+1)^2} - \frac{c^2}{N+1}\right)\right]^{-1}\frac{N}{N+1}}$$

$$= c\sigma\sqrt{\frac{N(N+1)}{N^2 - c^2(N+1)}}$$

   The term in parentheses is greater than 1, decreasing in N, and increasing in c. Notice that the contaminated value pulls $\bar{X}_{N+1}$ toward it and inflates $\hat{\sigma}^2_{N+1}$. These effects jointly make it more difficult to detect the outlier, a result known as outlier masking. Put another way, if X_{N+1} was removed and the thresholds were based on the mean and variance of the other observations, the detection threshold would be lower.

3 (17 pts) Mixture MLEs

Let $Y_1 \sim N(\mu_1, \sigma^2)$ and $Y_2 \sim N(\mu_2, \sigma^2)$. We know that an observation X_i, $i = 1, \ldots, N$, comes from distribution 1 with probability p.

Using `R`, generate 100 observations with $\mu_1 = 0$, $\mu_2 = 5$, $\sigma = 1$, and $p = 0.5$.

1. (4 pts) Derive the likelihood for the simulated data.

Solution: First, perform the following reparameterizations:

$$p = \frac{1}{1+\exp(\lambda)}$$
$$\sigma^2 = \exp(\beta).$$

The log likelihood is

$$\sum_{i=1}^{N} \log\left(\frac{1}{1+\exp(\lambda)}\frac{1}{\sqrt{\exp(\beta)}}\phi\left(\frac{x_i-\mu_1}{\sqrt{\exp(\beta)}}\right) + \frac{\exp(\lambda)}{1+\exp(\lambda)}\frac{1}{\sqrt{\exp(\beta)}}\phi\left(\frac{x_i-\mu_2}{\sqrt{\exp(\beta)}}\right)\right).$$

2. (4 pts) Code the likelihood function using `R`.
   Solution: We can code it as:

```
MixtureLL <-
function(theta, Y){
    mu_1  <- theta[1]
    mu_2  <- theta[2]
    sigma <-
exp(theta[3])
    p_1 <- 1 / (1
+ exp(theta[4]))
    -sum(log(p_1     *
dnorm(Y, mean = mu_1, sd
= sigma) +
             (1-p_1) *
dnorm(Y, mean = mu_2, sd
= sigma)))
}
```

3. (1 pt) Find the MLE values by setting the starting values for the means to 0 and the starting values for the other parameters to their true values.
   Solution: We find:

```
set.seed(271015)
X <- rbinom(100,
1, 0.5)
```

```
Y <- rnorm(100,
0 + 5*X)
optim(c(0,
0, 0, 0), MixtureLL,
Y = Y)

$par
[1]  2.489  1.535  0.961 -2.586

$value
[1] 238

$counts
function gradient
     181       NA

$convergence
[1] 0

$message
NULL
```

4. (1 pt) Find the MLE values by setting the starting values for p to near 0 and the starting values for the other parameters to their true values.
Solution: We have:

```
optim(c(0,
5, 10, 0), MixtureLL,
Y = Y)

$par
[1] 2.959 2.424 0.966 8.322

$value
[1] 238
```

```
$counts
function gradient
     203       NA

$convergence
[1] 0

$message
NULL
```

5. (1 pt) Find the MLE values by setting the starting values to the true values of the parameters.
Solution: We have:

```
optim(c(0,
5, 0, 0), MixtureLL,
Y = Y)

$par
[1] -0.05065  4.88844 -0.11167  0.00336

$value
[1] 200

$counts
function gradient
     153       NA

$convergence
[1] 0

$message
NULL
```

6. (6 pts) Using the results that maximize the likelihood, find the variance of the MLE.
Solution: We use the delta method to calculate the variance for the reparameteriza-

tions:

$$\text{Var}\left(\hat{p}\right) = \left[\frac{\exp(\lambda)}{\left(1+\exp(\lambda)\right)^2}\right]^2 \text{Var}\left(\hat{\lambda}\right)$$
$$\text{Var}\left(\hat{\sigma}^2\right) = \left[\exp(\beta)\right]^2 \text{Var}\left(\hat{\beta}\right).$$

Next, we recognize that maximizing the likelihood is equal to minimizing the negative of the log likelihood. The negative log likelihood is, unsurprisingly, found when we set the starting values to the true parameter values. Hence, we have:

```
mle_mix <-
optim(c(0,
5, 0, 0), MixtureLL,
Y = Y, hessian = TRUE)
diag(solve(mle_mix$hessian))
*
  c(1, 1,
exp(2*mle_mix$par[3]),
    (exp(mle_mix$par[4])/(1
+
exp(mle_mix$par[4])^2))^2)

[1] 0.01622 0.01631 0.00411 0.01004
```

4 (25 pts) Distribution of income

Vilfredo Pareto claimed that the distribution of household wealth could be described using what has come to be known as the Pareto distribution. In this question, we apply this distribution to household income.

For an income of $\$x$, the Pareto distribution implies:

$$F(x) = 1 - \left(\frac{1}{x}\right)^{\alpha}.$$

According to the U.S. Census Bureau, the distribution of income in 2014 is given in Table 1. These results are based on a survey of 99,461 households.

1. (8 pts) Derive (perhaps to a constant factor) the log likelihood using these data.

Table 1: Distribution of income, 2014

Income range	Percent in range
$0 - \$21,432$	20
$\$21,432 - \$41,186$	20
$\$41,186 - \$68,212$	20
$\$68,212 - \$112,262$	20
$\$112,262 - \$206,568$	15
$\$206,268+$	5

Solution: Create variables a_i, b_i, c_i, d_i, e_i, and f_i that indicate whether the person is in the respective six income buckets in Table 1. Consider the probablility of being in the first income quantile as specified by the Pareto distribution:

$$\Pr(a_i = 1) = \Pr(X < 21,432) = 1 - \left(\frac{1}{21,432}\right)^{\alpha}.$$

The probablity of being in the second quantile is:

$$\begin{aligned}\Pr(b_i = 1) = \Pr(21,432 < X < 41,186) &= \left[1 - \left(\frac{1}{41,186}\right)^{\alpha}\right] - \left[1 - \left(\frac{1}{21,432}\right)^{\alpha}\right] \\ &= \left(\frac{1}{21,432}\right)^{\alpha} - \left(\frac{1}{41,186}\right)^{\alpha}.\end{aligned}$$

The probabilities of being in the other buckets are calculated analogously.

Then, the likelihood is:

$$\prod_{i=1}^{N} \Pr(a_i = 1)^{a_i} \Pr(b_i = 1)^{b_i} \Pr(c_i = 1)^{c_i} \Pr(d_i = 1)^{d_i} \Pr(e_i = 1)^{e_i} \Pr(f_i = 1)^{f_i}.$$

The log likelihood is:

$$\begin{aligned}&\sum_{i=1}^{N} a_i \log\left(1 - \left(\frac{1}{21,432}\right)^{\alpha}\right) + \sum_{i=1}^{N} b_i \log\left(\left(\frac{1}{21,432}\right)^{\alpha} - \left(\frac{1}{41,186}\right)^{\alpha}\right) + \\ &\quad \sum_{i=1}^{N} c_i \log\left(\left(\frac{1}{41,186}\right)^{\alpha} - \left(\frac{1}{68,212}\right)^{\alpha}\right) + \sum_{i=1}^{N} d_i \log\left(\left(\frac{1}{68,212}\right)^{\alpha} - \left(\frac{1}{112,262}\right)^{\alpha}\right) + \\ &\quad \sum_{i=1}^{N} e_i \log\left(\left(\frac{1}{112,262}\right)^{\alpha} - \left(\frac{1}{206,568}\right)^{\alpha}\right) - \sum_{i=1}^{N} f_i \alpha \log(206,268)\end{aligned}$$

Notice that the only term inside the sums that varies across individuals is the indicator term. Hence, we're really summing over that and multiplying by the probabilities of being in each bucket. If the sum of number of people in each bucket are divied by the number of observations, we get probabilities. Hence, we get:

$$0.2\log\left(1-\left(\frac{1}{21,432}\right)^{\alpha}\right)+0.2\log\left(\left(\frac{1}{21,432}\right)^{\alpha}-\left(\frac{1}{41,186}\right)^{\alpha}\right)+$$
$$0.2\log\left(\left(\frac{1}{41,186}\right)^{\alpha}-\left(\frac{1}{68,212}\right)^{\alpha}\right)+0.2\log\left(\left(\frac{1}{68,212}\right)^{\alpha}-\left(\frac{1}{112,262}\right)^{\alpha}\right)+$$
$$0.15\log\left(\left(\frac{1}{112,262}\right)^{\alpha}-\left(\frac{1}{206,568}\right)^{\alpha}\right)-0.05\alpha\log(206,268)$$

2. (3 pts) Derive the score using these data.
Solution: Yeah, let's not do that.

3. (4 pts) Code the log likelihood and score functions using `R`.
Solution: We have:

```
ParetoLL <-
function(alpha){
  val <- 0.2  *
log(1                      -
(1 /
21432)^alpha) +
         0.2  * log((1
/  21432)^alpha -
(1 /
41186)^alpha) +
         0.2  * log((1
/  41186)^alpha -
(1 /
68212)^alpha) +
         0.2  * log((1
/  68212)^alpha -
(1 /
112262)^alpha) +
         0.15 * log((1
/ 112262)^alpha -
(1 /
```

```
206568)^alpha) -
         0.05 * alpha *
log(206268)
  val <- - val
}
```

4. (2 pts) Find the MLE for α.
   Solution: We find:

```
mle_pareto <- optim(1,
ParetoLL)
mle_pareto

$par
[1] 0.0975

$value
[1] 3.15

$counts
function gradient
      36       NA

$convergence
[1] 0

$message
NULL
```

5. (4 pts) Using each of the six income buckets in Table 1 separately, give a MOM estimator for α.
   Solution: Here, we find the α such that the parametric probability for each bucket is equal to the observed probability in the bucket. These results are given in Table 2.

```
CalculateMoment <-
function(alpha, L,
U, P){
  moment <- (1 /
L)^alpha - (1
/ U)^alpha
  abs(P - moment)
}
mom1 <-
optim(mle_pareto$par, CalculateMoment,
L =      1, U =
21432, P = 0.2)
mom2 <-
optim(mle_pareto$par, CalculateMoment,
L =  21432, U =
41186, P = 0.2)
mom3 <-
optim(mle_pareto$par, CalculateMoment,
L =  41186, U =
68212, P = 0.2)
mom4 <-
optim(mle_pareto$par, CalculateMoment,
L =  68212, U =
112262, P = 0.2)
mom5 <-
optim(mle_pareto$par, CalculateMoment,
L = 112262, U =
206568, P = 0.15)
mom6 <-
optim(mle_pareto$par, CalculateMoment,
L = 206568, U =
1e9, P = 0.05)
```

6. (4 pts) How do all these estimates compare to one another? What conclusion do you draw?

   Solution: The two tails of the distribution give different MOM estimates than the

Table 2: ML and MOM estimates of α

Approach	Estimate
ML	0.097
MOM 1	0.022
MOM 2	0.097
MOM 3	0.092
MOM 4	0.088
MOM 5	0.084
MOM 6	0.233

ML estimates, but the other buckets provide similar results. Hence, we may conclude the the Pareto distribution reasonably matches the middle of the income distribution, but does poorly in the tails.

5 (25 pts) EL Chupacabra

A disadvantage of maximum likelihood estimation is that a parametric density must be specified. Consider the empirical density for X_i, ($i = 1, \ldots, N$ unique):

$$f(x_i) = \pi_i$$

for $0 \leq \pi_i \leq 1$ and $\sum_i \pi_i = 1$.

1. (3 pts) What is the empirical likelihood (*i.e.*, the likelihood based upon the empirical density)?
   Solution: The empirical log likelihood is

   $$\ell_N(x_i) = \sum_{i=1}^{N} \log(\pi_i).$$

2. (5 pts) Derive the MLE for $\{\pi_i\}$.

Solution: We solve the maximization problem:

$$\max_{\{\pi_i\},\lambda} \sum_{i=1}^{N} \log(\pi_i) - \lambda \left(\sum_{i=1}^{N} \pi_i - 1 \right)$$
$$\Rightarrow \frac{1}{\pi_i} - \lambda = 0 \quad \forall\, i$$
$$\pi_i = \frac{1}{\lambda} \quad \forall\, i$$
$$\Rightarrow \pi_i = \frac{1}{N} \quad \forall\, i$$

Now, consider maximizing the log empirical likelihood subject to the moment conditions:

$$\sum_{i=1}^{N} \pi_i m(x_i; \theta) = 0,$$

where $\theta \in \mathbb{R}^K$ and $m(\cdot;\cdot) \in \mathbb{R}^R$, $R \geq K$.

3. First, maximize over $\{\pi_i\}$ and the Lagrange multipliers, holding θ fixed.

   (a) (5 pts) What are the MLE for $\{\pi_i\}$?

   Solution: We solve the maximization problem:

$$\max_{\{\pi_i\},\lambda,\gamma} \sum_{i=1}^{N} \log(\pi_i) - \lambda \left(\sum_{i=1}^{N} \pi_i - 1 \right) - \gamma' \sum_{i=1}^{N} \pi_i m(x_i; \theta)$$
$$\Rightarrow \frac{1}{\pi_i} - \lambda - \gamma' m(x_i; \theta) = 0 \quad \forall\, i$$
$$\pi_i = \frac{1}{\lambda + \gamma' m(x_i; \theta)}$$
$$\pi_i = \frac{1}{N + \gamma' m(x_i; \theta)}$$

   Note that γ is a vector, as $m(\cdot;\cdot)$ is a vector.

   (b) (4 pts) What are the equations for the Lagrange multipliers?

   Solution: For λ, take the score for the probabilities, multiply by π_i, then sum across i:

$$\begin{aligned}
\sum_i \pi_i \left[\frac{1}{\pi_i} - \lambda - \gamma' m(x_i;\theta)\right] &= 0 \\
\sum_i \left[1 - \pi_i\lambda - \gamma'\pi_i m(x_i;\theta)\right] &= 0 \\
N - \lambda &= 0 \\
\lambda &= N
\end{aligned}$$

We plug this result in to part (a) above to get that result. This value is the same as for the maximization problem without the moment conditions.

We find the γ that solves:

$$\sum_{i=1}^{N} \pi_i = \sum_{i=1}^{N} \frac{1}{N + \gamma' m(x_i;\theta)} = 1$$

(c) (4 pts) How do the MLE for $\{\pi_i\}$ under the constrained optimization compare to the unconstrained MLEs? Provide an intuitive explanation.

Solution: The denominator is no longer simply N; instead, it is adjusted by a certain amount. This adjustment is used to ensure that the moment conditions hold. Notice that, unlike the maximization problem without the moment constraints, the observed values are no longer equally weighted.

(d) (2 pts) What is the purpose of adding the moment constraints?

Solution: The moment conditions introduce features of the distribution (*i.e.*, θ) that we are interested in estimating. Otherwise, we just get an estimate of the density, which may not be of specific interest.

4. (2 pts) Now, set up an unconstrained maximization problem for θ using a likelihood based on the MLEs for $\{\pi_i\}$ that you found in 3(a). (You do not need to solve it.)

Solution: We just maximize the sum of the log of the probabilities found in 3(a):

$$\max_\theta \sum_{i=1}^{N} \log\left(N + \gamma' m(x_i;\theta)\right) \quad \text{s.t.} \quad \sum_{i=1}^{N} \frac{1}{N + \gamma' m(x_i;\theta)} = 1.$$

This is known as the *empirical likelihood* estimation approach and is a special case of the generalized method of moments.