eman ta zabal zazu

Universidad del País Vasco Euskal Herriko Unibertsitatea

Activity 2

1 Synopsis.

What this activity is about. We have seen that seemingly natural candidates to estimate a location parameter can be a total failure, if the distribution does not have moments (Cauchy) or fails to verify regularity conditions (case of the $U(0, \theta)$, discussed in class).
We have seen also that a "cure" for the non-convergence of $\overline{X}_n$ to the location parameter d in the Cauchy distribution,

$$f(x) = \frac{1}{\pi} \frac{1}{1 + (x - d)^2} \qquad (-\infty < x < \infty)$$

is to use a *trimmed mean*, discarding far away observations. In the most extreme case of trimming, we only leave the central observation —the median— as an estimate of d.
You will explore now a similar situation, that of the Laplace distribution $L(\mu, \sigma)$, with density

$$f_X(x) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma} \qquad (-\infty < x < \infty) \tag{1}$$

Notice that it is similar to the normal, but the exponent of e is such that the density decays at a lower rate than that of the normal density, so there are heavy tails. This suggests that the same issues we found with the Cauchy will have to be faced here.
We recommend that you read once the problems, then read through Section 3 and finally come back to do the problems.

What you need. You need to be fully acquainted with the content of previous seminars and practice assignment. You will also need access to a computer equiped with R.

2 Problems

Consider a Laplace distribution, $L(\mu, \sigma = 1)$.

1. What is the moment estimator for μ?

2. Following the directions in Section 3, compute the mean and variance of $\overline{X}_n$. Is $\overline{X}_n$ unbiased as an estimator of μ?

3. What is the MLE of μ?

4. Making use of the factorization theorem, search for sufficient statistics for μ in the distribution $L(\mu, 1)$ ($\sigma = 1$, assumed known). Are there any different from the trivial $(X_{(1)}, \ldots, X_{(n)})$?

5. Look at the plot of the log likelihood associated to a sample, for values of μ from +1 to +5 (in Section 3, item 7). What do you observe? Is the log likelihood differentiable as a function of μ? Explain why or why not.

6. Can you compute the Cramér-Rao lower bound for unbiased estimators in this problem?

7. Consider now the average $\overline{X}_n$ and the sample median $\tilde{X}_n$, both based on n observations[1], as estimators of μ. You have computed the moments of $\overline{X}_n$ above. Now, a standard result (given in Section 3, item 8) gives the distribution of the median, $\tilde{X}_n$, for large samples. If we were sampling a normal distribution, $N(\mu, 1)$, which of $\overline{X}_n$ and $\tilde{X}_n$ would be more efficient as an estimator of μ when n is large?

8. Repeat the preceding analysis if rather than a $N(\mu, 1)$ we are sampling from a $L(\mu, 1)$. What are now the large sample variances of $\overline{X}_n$ and $\tilde{X}_n$? What is the *relative* efficiency of $\tilde{X}_n$ as compared to $\overline{X}_n$?

9. In the case of the $L(\mu, 1)$ is any of $\overline{X}_n$ or $\tilde{X}_n$ consistent? Both? If both converge in probability to the true value of μ, which does so faster?

10. We rarely can tell whether our sample comes from a normal or a Laplace. You have seen that depending on which distribution you sample, a different estimator is best. It would be nice to have an "all around" estimator, that would work perhaps not optimally, but at least fairly well in any situation. A trimmed mean, which is sort of half way between the ordinary mean and the median, seems like a good candidate.

    Perform a small simulation to compare $\overline{X}_n$, $\tilde{X}_n$ and an α trimmed mean $\overline{X}_n^{\alpha}$ (you remove $n\alpha/2$ observations on each side; for instance, for a 0.20 trimmed mean with $n = 100$ observations you would discard the ten smallest and ten largest observations.)

    We suggest you to keep your simulation small in size. For instance, you may try ten sample sizes between $n = 50$ and $n = 10000$ and for each sample size:

    (a) Generate $k = 500$ samples from a normal $N(\mu, 1)$ of your choice.

    (b) For each of the k samples,

       i. Compute $\overline{X}_n$, $\tilde{X}_n$ and $\overline{X}_n^{\alpha}$ with $\alpha = 0.20$ trimming.

       ii. Compute the square error, $(\overline{X}_n - mu)^2$, etc.

    (c) Average for each estimator the square errors over the k replications. That should be a good approximation of the MSE of each estimator. Store that.

    Then, do it for the $L(\mu, 1)$: you can reuse your code replacing at a single spot `rnorm` by `rlaplace`.

    The final output should be a table listing MSE (close to variance, as all estimators are at least asymptotically unbiased) for each combination of n, distribution and estimator. The values for $n > 1000$ should be close to the theoretical values that you have produced in your previous answers.

[1] We will assume n odd, so the sample median is unique.

3 Hints, details, comments

1. The moments of the Laplace distribution can be obtained directly. The mean is given by:

$$E[X] = \int_{-\infty}^{\infty} x \frac{1}{2\sigma} e^{-|x-\mu|/\sigma} dx.$$

2. If we standardize the variable defining $Y = (X - \mu)/\sigma$, then

$$f_Y(y) = f_X(\mu + \sigma y) \left| \frac{d}{dy}(\mu + \sigma y) \right| = \frac{1}{2} e^{-|y|}$$

$$E[X] = E[\mu + \sigma Y] = \mu + \sigma E[Y]. \tag{2}$$

3. Now,

$$E[Y] = \int_{-\infty}^{\infty} y \frac{1}{2} e^{-|y|} dy = \int_{-\infty}^{0} y \frac{1}{2} e^{-|y|} dy + \int_{0}^{\infty} y \frac{1}{2} e^{-|y|} dy; \tag{3}$$

this would be trivially be zero (because the integrand is an odd function of y, and the integral over $[0, \infty)$ would thus cancel the integral over $(-\infty, 0]$) *provided* each of these integrals is finite; you cannot perform $\infty - \infty$.

Finiteness is easy to show; if we take the second integral in (3) we have:

$$\int_{0}^{\infty} y \frac{1}{2} e^{-|y|} dy = \frac{1}{2} \int_{0}^{\infty} y e^{-y} dy$$

and the last integral is 1; we can recognize the integrand as the density of a $\gamma(a = 1, r = 2)$, which, like all densities, integrate to 1. Therefore, $E[Y] = 0$ and replacing that value in (2) we have $E[X] = \mu$.

4. You can use the same trick to compute $E[Y^2]$. You will find that

$$E[Y^2] = \frac{1}{2} \int_{-\infty}^{\infty} y^2 e^{-|y|} dy = \int_{0}^{\infty} y^2 e^{-|y|} dy; \tag{4}$$

multiplying inside and outside the integral by, respectively, c and $1/c$, you should be able to recognize the integral of a $\gamma(a = 1, r = 3)$. Once you have $E[Y^2]$, computing the variance of X is easy:

$$\text{Var}[X] = E[(X - \mu)^2] = \sigma^2 E[Y^2].$$

5. If you want to minimize $\sum_i |x_i - \mu|$ you should not try to take derivatives and equate to zero; that will not work. However, it is easy to realize that the minimizing μ is one median of the set $x_1, \ldots, x_n$ (see note[2]).

6. You do not have in R base the usual set of `d,p,q,r` functions for the Laplace distribution. However, they are available in package `rmutil`. Install it in your machine if you do not have it. Examples of use:

```
> require(rmutil)
> dlaplace(3.4, m=0.2, s=1)          # density

[1] 0.0203811

> plaplace(3.4, m=0.2, s=1)          # distribution function

[1] 0.9796189
```

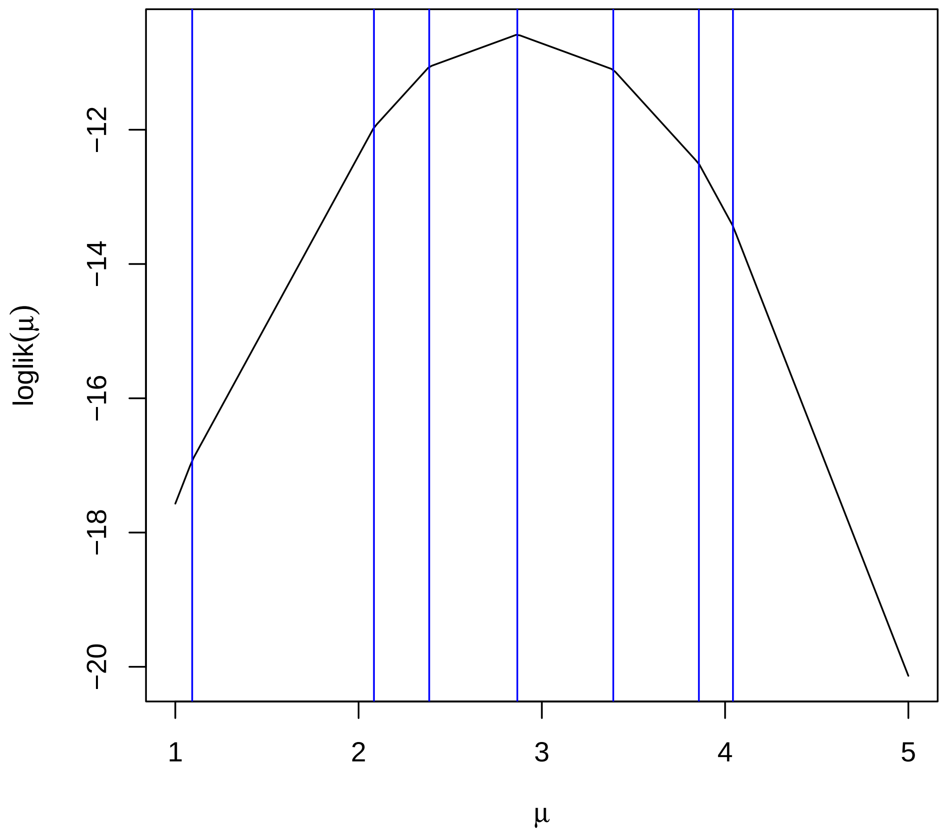
[2] Why? As long as μ has more observations to its right than to its left, $\sum_i |x_i - \mu|$ can be reduced by moving μ a little to the right: what we gain is less than what we loose. Incidentally, you can think of this in terms of Hotelling's spatial competition model, which you may have encountered in Microeconomics.

```
> qlaplace(0.9796189, m=0.2, s=1)    # quantile function; check that it matches

[1] 3.4

>                                    # previous result
> rlaplace(6, m=0.2, s=1)            # random number generation

[1]  0.6770826  0.9339897  0.6545274  0.3248643  0.4826118
[6] -2.3535784
```

7. Observe the following chunk of code. It generates a sample of $n = 7$ observations from the Laplace distribution $L(\mu = 3, \sigma = 1)$, defines a function (`loglike`) that computes the log likelihood of a single value of μ and then uses that function in a loop to compute the log likelihood for values of μ going from $+1$ to $+5$. A plot is finally made, and the position of each observation marked with a blue vertical line.

   Satisfy yourself that you know what each instruction does; experiment if you do not understand something, executing the code one line at a time, and ask your teacher if need be.

```
> set.seed(12348)                         # for reproducibility
> require(rmutil)                         # needed for function rlaplace below
> n <-7 ; mu <- 3 ; sigma <- 1            # changeable values here
> sample <- rlaplace(n, m=mu, s=sigma)    # sample generation
> loglike <- function(mu) {
      logver <- n * log(1/2*sigma) - sum(abs(sample-mu)/sigma)
      return(logver)
      }
> mus     <- seq(from=1, to=5, by=0.02)   # values of mu to compute log likelihood
> logver  <- 0 * mus                      # placeholder for log likelihood values
> for (i in 1:length(mus)) {              # here we go; this loop makes the actual
      logver[i] <- loglike( mus[i] )      # computation.
      }
> plot(mus, logver,                       # we draw de log likelihood
       xlab=expression(mu),
       ylab=expression(loglik(mu)),
       type="l")
> abline(v=sample, col="blue")            # plot the location of the sample values
> # mm <- median(sample)                  # if desired, we could mark the position
> # abline(v=mm, col="red")               # of the median as well
```



8. Under some conditions (symmetric distribution about μ, which is the unique median, continuous and positive density at $x = \mu$), then for large sample size n the median is approximately distributed as

$$\tilde{X}_n \sim N\left(\mu, \frac{1}{4n(f(\mu))^2}\right)$$

(a proof can be found in advanced books, such as [1], p. 354). For instance, if our observations come from a $N(5, 1)$, $\mu = 5$, $f(5) = \frac{1}{\sqrt{2\pi}} e^{-(5-5)^2/2} = 0.3989423$. The variance of $\tilde{X}_n$ would then be

$$\frac{1}{4n \times 0.3989423^2} = 1.570948/n.$$

References

[1] E. L. Lehmann. *Theory of Point Estimation.* Wiley, New York, 1983.