University of California, Los Angeles
Department of Statistics

Statistics 100B **Instructor: Nicolas Christou**

Method of maximum likelihood

Suppose $x_1, x_2, \cdots, x_n$ is a random sample of size n from a distribution that has parameter θ. The joint probability density of these n random variables is

$$f(x_1, x_2, \cdots, x_n; \theta)$$

We also refer to this function as the *likelihood function* and it is denoted with L. In this function the parameter θ is unknown and it will be estimated with the method of maximum likelihood. In principle, the method of maximum likelihood consists of selecting the value of θ that maximizes the likelihood function (the value of θ that makes the observed data more likely).

Since $x_1, x_2, \cdots, x_n$ are independent the likelihood function can be expressed as the product of the marginal densities:

$$L = f(x_1, x_2, \cdots, x_n; \theta) = f(x_1; \theta) \times f(x_2; \theta) \times \cdots \times f(x_n; \theta)$$

We will maximize this function w.r.t. θ. It is often easier to maximize the *log likelihood* function w.r.t. θ. Therefore, we will take the derivative of the log likelihood function w.r.t. θ, set it equal to zero and solve for θ. The result will be denoted with $\hat{\theta}$ and we refer to it as the *mle* of the parameter θ.

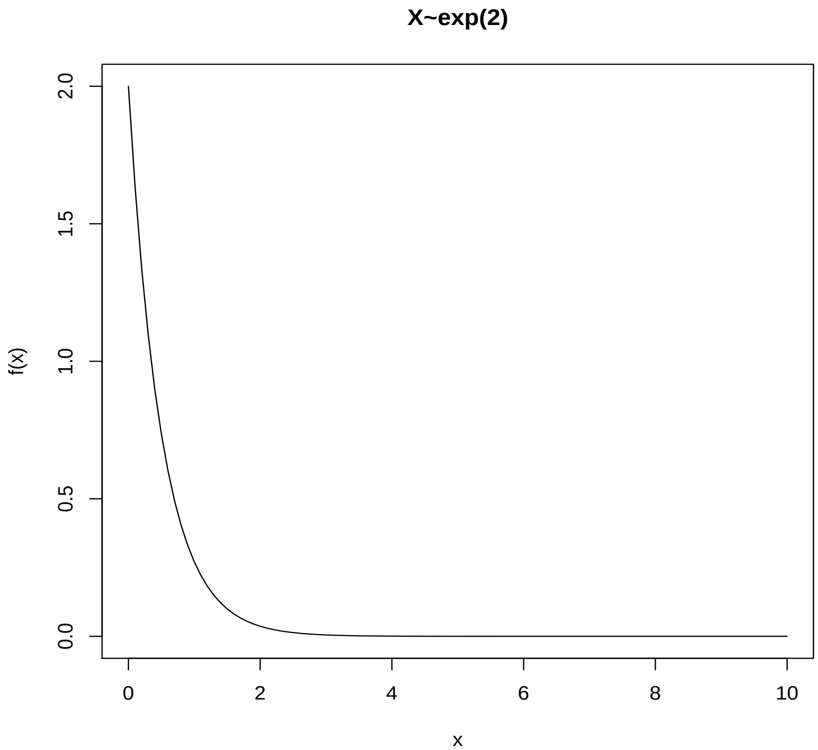
Example:
Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from an exponential distribution with parameter λ. Find the mle of λ.

Example:

Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Find the mle of μ and σ^2.

Method of maximum likelihood - An empirical investigation

We will estimate the parameter λ of the exponential distribution with the method of maximum likelihood. Let $X \sim exp(2)$ (see figure below).



Let's pretend that λ is unknown. From this distribution we will select a random sample of size $n = 100$ (see observations on the next page). This sample gave $\sum_{i=1}^{100} x_i = 49.86463$ and sample mean $\bar{x} = 0.4986463$. Therefore, the method of maximum likelihood estimate of λ is: $\hat{\lambda} = \frac{1}{\bar{x}} = \frac{1}{0.4986463} = 2.005429$.

For different values of the parameter λ we compute the log-likelihood function as follows:

$$ln(L) = nln(\lambda) - \lambda \sum_{i=1}^{100} x_i$$

These calculations are shown on the next page. We then plot the values of the log likelihood function against λ and we observe that the maximum occurs at the value of $\hat{\lambda} = 2.005429$ that was computed above.

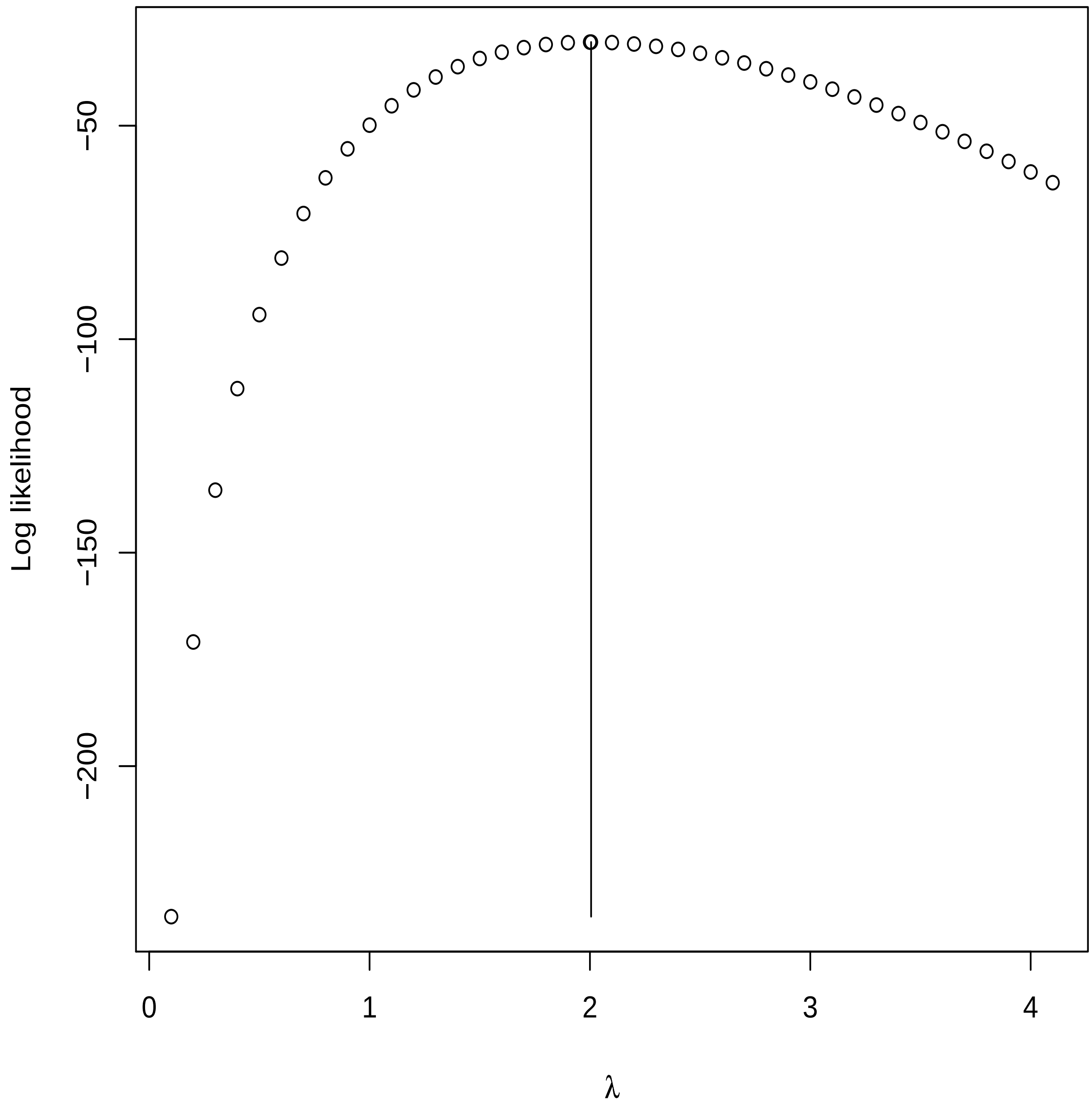
Observations of a random sample of size $n = 100$ from exponential distribution with $\lambda = 2$:

```
  [1] 1.695824351 0.066702402 0.674994950 0.736106579 1.161993229
  [6] 0.296223724 0.043937990 0.508988160 0.294233621 0.024084168
 [11] 0.150176375 0.396972182 0.095883055 0.387135421 0.248432954
 [16] 0.661809923 0.142542189 0.171455182 1.212420122 0.180640216
 [21] 0.009212488 0.160395423 0.188922063 0.884223028 0.240872947
 [26] 0.033885428 0.080997465 0.318024634 0.410324188 0.502538879
 [31] 0.422821270 0.329996007 0.446404769 0.522652992 0.154471200
 [36] 0.064116746 0.268321347 0.263458486 0.581443048 1.031375370
 [41] 0.203961618 2.562959307 0.073292671 1.025867874 0.173630370
 [46] 0.263878938 0.171617840 0.028656404 1.961520632 0.242559879
 [51] 0.491987590 0.410541936 0.500918018 0.322782228 1.497851781
 [56] 0.157720428 0.629583415 0.652147642 0.135310800 1.936474929
 [61] 0.181363227 0.227498170 1.490756486 0.334677184 0.368089615
 [66] 0.272378459 0.525470783 0.476837360 0.224213297 0.171204443
 [71] 0.119797853 0.716180556 0.111337474 0.376437023 0.588020059
 [76] 0.156395280 0.135622347 0.067554610 1.745086826 1.661906995
 [81] 0.023611775 0.080141754 0.089054515 0.004390821 1.183269692
 [86] 0.199572674 1.043889988 1.136122111 0.545845778 0.234890293
 [91] 0.558763671 0.196966494 0.692430989 0.342892071 0.369322342
 [96] 0.671608332 0.254633346 0.076204614 0.157962865 2.543944322
```

Values of the log likelihood function for different λ:

```
       lambda        lnL
 [1,] 2.00543  -30.41417
 [2,] 0.10000 -235.24497
 [3,] 0.20000 -170.91672
 [4,] 0.30000 -135.35667
 [5,] 0.40000 -111.57492
 [6,] 0.50000  -94.24703
 [7,] 0.60000  -81.00134
 [8,] 0.70000  -70.57273
 [9,] 0.80000  -62.20606
[10,] 0.90000  -55.41421
[11,] 1.00000  -49.86463
[12,] 1.10000  -45.32007
[13,] 1.20000  -41.60539
[14,] 1.30000  -38.58759
[15,] 1.40000  -36.16325
[16,] 1.50000  -34.25043
[17,] 1.60000  -32.78304
[18,] 1.70000  -31.70704
[19,] 1.80000  -30.97766
[20,] 1.90000  -30.55740
[21,] 2.00000  -30.41453
[22,] 2.10000  -30.52198
[23,] 2.20000  -30.85644
[24,] 2.30000  -31.39773
[25,] 2.40000  -32.12823
[26,] 2.50000  -33.03249
[27,] 2.60000  -34.09688
[28,] 2.70000  -35.30931
[29,] 2.80000  -36.65901
[30,] 2.90000  -38.13634
[31,] 3.00000  -39.73265
[32,] 3.10000  -41.44013
[33,] 3.20000  -43.25172
[34,] 3.30000  -45.16102
[35,] 3.40000  -47.16218
[36,] 3.50000  -49.24989
[37,] 3.60000  -51.41927
[38,] 3.70000  -53.66583
[39,] 3.80000  -55.98547
[40,] 3.90000  -58.37438
[41,] 4.00000  -60.82907
[42,] 4.10000  -63.34627
```

Plot of the log likelihood function against λ:



Properties of estimators, method of maximum likelihood - Examples

Example 1:
Let X follow the uniform distribution on the interval $(0, \theta)$. Find the mle of θ.

Example 2:
Let $f(x; \theta) = \theta x^{\theta - 1}, \quad 0 < x < 1, \quad 0 < \theta < \infty$. Find the mle of θ.

Example 3:
Let $f(x; \theta) = e^{-(x-\theta)}, \quad 0 < x < \infty, \quad \theta < x$. Find the mle of θ.

Example 4:
Suppose that $X_1, \cdots, X_m$ representing yields per acre for corn variety A, is a random sample from $N(\mu_1, \sigma)$. Also, $Y_1, \cdots, Y_n$ representing yields for corn variety B, is a random sample from $N(\mu_2, \sigma)$. If the two samples are independent, find the maximum likelihood estimate for the common variance σ^2. Assume that μ_1 and μ_2 are unknown.

Example 5:
Let $X_1, \cdots, X_n$ denote a random sample from the probability density function $f(x; \theta) = (\theta + 1)x^{\theta}, \quad 0 < x < 1, \quad \theta > -1$. Find the mle of θ.

Example 6:
In a random sample of 100 men 25 are Democtrats, and in a random sample of 100 women 30 are Democrats. The two samples are independent. Assume that p_M is the true proportion of Democrats among all men, and p_W is the true proportion of Democrats among alll women. Suppose that $p_M = p_w = p$. Find the mle of the common proportion p.

Problem 7
Let $X_1, X_2, \cdots, X_n$ be an i.i.d. random sample from $N(\mu, \sigma)$.

a. Which of the following estimates is unbiased? Show all your work.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}, \quad S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

b. Which of the estimates of part (a) has the smaller MSE?

Problem 8
Let $X_1, X_2, \cdots, X_n$ be an i.i.d. random sample from a normal population with mean zero and unknown variance σ^2.

a. Find the maximum likelihood estimate of σ^2.

b. Show that the estimate of part (a) is unbiased estimator of σ^2.

c. Find the variance of the estimate of part (a). Is it consistent?

d. Show that the variance of the estimate of part (a) is equal to the Cramer-Rao lower bound.

Problem 9
Let $X_1, X_2, \cdots, X_n$ denote an i.i.d. random sample from the exponential distribution with mean $\frac{1}{\lambda}$.

a. Derive the maximum likelihood estimate of λ.

b. Find the Cramer-Rao lower bound of the estimator of λ.

c. What is the asymptotic distribution of $\hat{\lambda}$?

Problem 10
Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed random variables from a Poisson distribution with parameter λ. We know that the maximum likelihood estimate of λ is $\hat{\lambda} = \bar{x}$.

a. Find the variance of $\hat{\lambda}$.

b. Is $\hat{\lambda}$ an MVUE?

c. Is $\hat{\lambda}$ a consistent estimator of λ?

Problem 11
Suppose that two independent random samples of n_1 and n_2 observations are selected from two normal populations. Further, assume that the populations possess a common variance σ^2 which is unknown. Let the sample variances be S_1^2 and S_2^2 for which $E(S_1^2) = \sigma^2$ and $E(S_2^2) = \sigma^2$.

a. Show that the pooled estimator of σ^2 that we derived in class below is unbiased.

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

b. Find the variance of S^2.

Problem 12
In a basket there are green and white marbles. You randomly select marbles with replacement until you see a green marble. You found the first green marble on the $10th$ trial. Then, your friend does the same. He randomly selects marbles until he obtains a green marble. His green marble was seen on the $15th$ trial. Use the method of maximum likelihood to find an estimate of p, the proportion of green marbles in the basket.

Theorem
Asymptotic efficiency of maximum likelihood estimates.
Why do maximum likelihood estimates have an asymptotic normal distribution? Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from a probability density function $f(x|\theta)$. Then if $\hat{\theta}$ is the MLE of θ the theorem states that $\hat{\theta} \sim N(\theta, \sqrt{\frac{1}{nI(\theta)}})$.

Proof
We will use Taylor series. This says that for a function h

$$h(y) \approx h(y_0) + h'(y_0)(y - y_0).$$

Start with the likelihood function $L = \Pi_{i=1}^n f(x_i|\theta)$. Then the log-likelihood is

$$ln(L) = \sum_{i=1}^n lnf(x_i|\theta).$$

Now obtain the derivative w.r.t. θ.

$$\frac{\partial}{\partial\theta} ln(L) = \sum_{i=1}^n \frac{\partial}{\partial\theta} lnf(x_i|\theta).$$

Now letting $\hat{\theta}$ be the MLE of θ we write this as a Taylor series about that $\hat{\theta}$:

$$\sum_{i=1}^n \frac{\partial}{\partial\theta} lnf(x_i|\theta) \approx \sum_{i=1}^n \frac{\partial}{\partial\theta} lnf(x_i|\hat{\theta}) + \left[\sum_{i=1}^n \frac{\partial^2}{\partial\theta^2} lnf(x_i|\hat{\theta})\right](\theta - \hat{\theta})$$

Now divide left and right by $\sqrt{n}$ to get:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial\theta} lnf(x_i|\theta) \approx \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial\theta} lnf(x_i|\hat{\theta}) + \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial^2}{\partial\theta^2} lnf(x_i|\hat{\theta})\right](\theta - \hat{\theta})$$

Note: The first term on the right hand side is zero (because this is what we do to find $\hat{\theta}$). Therefore, we have reduced the relationship to the following:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial\theta} lnf(x_i|\theta) \approx \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial^2}{\partial\theta^2} lnf(x_i|\hat{\theta})\right](\theta - \hat{\theta})$$

Examine the left hand side: This involves the sum of n independent, identically distributed things (Central Limit theorem). Each one of these "things" has mean zero and variance $I(\theta)$. Therefore the left hand side follows approximately $N(0, I(\theta))$. Why?

Therefore, the limiting distribution of the right hand side must also be $N(0, I(\theta)$, i.e.

$$\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial^2}{\partial\theta^2} lnf(x_i|\hat{\theta})\right](\theta - \hat{\theta}) \sim N(0, I(\theta)).$$

Or write it as (watch the $n's$ and the minus sign!):

$$\left[-\frac{1}{n} \sum_{i=1}^n \frac{\partial^2}{\partial\theta^2} lnf(x_i|\hat{\theta})\right]\sqrt{n}(\hat{\theta} - \theta) \sim N(0, I(\theta)).$$

The expression in the bracket converges to $I(\theta)$ (law of large numbers) and therefore we can express the previous expression as

$$I(\theta)\sqrt{n}(\hat{\theta} - \theta) \sim N(0, I(\theta)),$$

or

$$\hat{\theta} \sim N(\theta, \sqrt{\frac{1}{nI(\theta)}}).$$