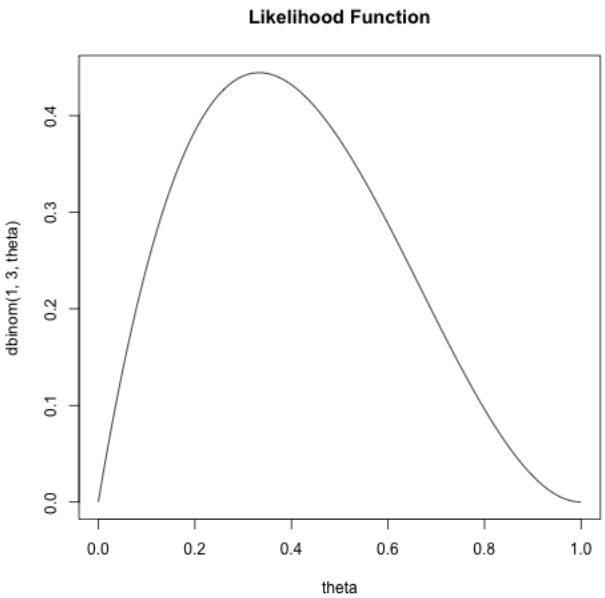
Maximum likelihood Estimate (MLE)

Let $f(x_1, \cdots, x_n; \theta)$ be the joint pdf/pmf of $X_1, \cdots, X_n$.
For fixed $x_1, \cdots, x_n$,
$L(\theta) = f(x_1, \cdots, x_n; \theta)$ (as a function of θ) is called the **likelihood function** .
$\log L(\theta)$ is called the loglikelihood function.
If $X_1, \cdots, X_n$ iid (independent and identically distributed) with pdf $f(x; \theta) > 0$, then
$L(\theta) = f(x_1; \theta) f(x_2; \theta) ... f(x_n; \theta)$ and
$\log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta)$.
e.g. Assume $X \sim binomial(3, \theta)$, then
$P(X = 1) = 3\theta(1 - \theta)^2 := L(\theta)$ is the likelihood function and
$LogL(\theta) = \log(3) + \log(\theta) + 2\log(1 - \theta)$ is the loglikelihood function.


Likelihood Function
dbinom(1, 3, theta)
0.0
0.1
0.2
0.3
0.4
0.0
0.2
0.4
0.6
0.8
1.0
theta

Likelihood function

More general given $X = x$,
$P(X = x) = \frac{3!}{x!(3-x)!}\theta^x(1-\theta)^{3-x} = L(\theta)$
If assume X_1, X_2 iid, and observe $X_1 = 1, X_2 = 2$, then
$P(X_1 = 1, X_2 = 2) = 3\theta(1-\theta)^2 * 3\theta^2(1-\theta) = 9\theta^3(1-\theta)^3 = L(\theta)$.
More general, given $X_1 = x_1, X_2 = x_2$, we can get
$L(\theta) = \binom{3}{x_1}\binom{3}{x_2}\theta^{x_1+x_2}(1-\theta)^{6-(x_1+x_2)}$.
Further, if we assume X_i iid $\sim \mathit{binomial}(k, \theta)$ and observe
$x_1, x_2, \cdots x_n$, then
$L(\theta) = (\prod_{i=1}^n \binom{k}{x_i})\theta^{\sum_{i=1}^n x_i}(1-\theta)^{kn-\sum x_i}$.

Note the product notation $\prod a_i = a_1 a_2 .. a_n$

MLE

For a given set of observations, $x_1, \cdots, x_n$, the **maximum likelihood estimate** of θ is a point $\hat{\theta} \in \Theta$, say $\hat{\theta} = h(x_1, \cdots, x_n)$ satisfying
$f(x_1, \cdots, x_n; \theta) = \max_{\theta \in \Theta} f(x_1, \cdots, x_n; \theta)$.
i.e.,$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$.
The MLE is the parameter point for which the observed sample is most likely.

Find MLE

Differentiating or direct maximization.
Assume $L(\theta)$ is twice differentiable in the interior of Θ, then $\hat{\theta}$ maximizes $L(\theta)$ if $\hat{\theta}$ is the unique extreme point in Θ satisfying $\frac{dL(\theta)}{d\theta}|_{\theta=\hat{\theta}} = 0$, $\frac{d^2L(\theta)}{d\theta^2}|_{\theta=\hat{\theta}} < 0$
$L(\theta)$ can be replaced by $\log L(\theta)$ if loglikelihood is well defined.

Examlpe: $X \sim bin(3, \theta)$, observe $X = 1$ Find MLE of θ.
$\log L(\theta) = \log(3) + \log(\theta) + 2\log(1-\theta)$,
Let $\frac{d \log L(\theta)}{d\theta} = \frac{1}{\theta} + 2\frac{-1}{1-\theta} = 0$,
solve for $\hat{\theta} = \frac{1}{3}$.
If observe $X = x$, then
$\log L(\theta) = \log(\binom{3}{x}) + x\log(\theta) + (3-x)\log(1-\theta)$,
by setting $\frac{d \log L(\theta)}{d\theta} = 0$,
we can obtain $\hat{\theta} = \frac{x}{3}$.

Find MLE

example: let $X_1, \cdots, X_n$ be a random sample from an exponential distribution $f(x;\theta) = \frac{1}{\theta}e^{-\frac{1}{\theta}x}$, find the MLE for θ.
Note the likelihood function is $L(\theta) = (\frac{1}{\theta})^n e^{-\frac{1}{\theta}\sum x_i}$,
$logL(\theta) = -n\log(\theta) - \frac{1}{\theta}\sum x_i$
Let $\frac{d\log L(\theta)}{d\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum x_i = 0$
We can obtain $\hat{\theta} = \frac{\sum X_i}{n} = \bar{X}$.

Multiple parameters

Differentiate $L(\theta)$ with respect to each θ_i to find MLE.
$\frac{\partial L(\theta)}{\partial \theta_i} = 0, i = 1, \cdots, k$.
Example: $X_1, \cdots, X_n$ iid $\sim N(\mu, \sigma^2)$. Find MLE of μ and σ^2.

solutions: $\hat{\mu} = \bar{X}_n, \hat{\sigma}^2 = \frac{1}{n} \sum_i (X_i - \bar{X}_n)^2$.
Derivation is not required here. You only need to know you can set partial derivative with respect to each parameter to 0 to find MLE for multiple parameters.

Direct maximization

Sometimes we need to use direct maximization to find MLE when taking derivative does not work.
example: $X_1, \cdots, X_n$ iid $\sim$ Uniform $(0, \theta)$.
Find MLE of θ. Note θ here is a range parameter.
$\hat{\theta} = X_{(n)}$ where $X_{(n)} = \max(X_1, X_2, \cdots X_n)$.
To see why, assume we abserve $x_1 = 1.1, x_2 = 1.8$ for $n = 2$.
$L(\theta) = (\frac{1}{\theta})^2$ if $\theta \geq 1.8$
Note the smaller the value for θ, the larger $L(\theta)$, but θ has to be at least 1.8, so $\hat{\theta} = 1.8$.

exercise

1. One observation is taken on a discrete random variable X with pmf $f(x;\theta)$, where $\theta \in \{1,2,3\}$. Find the MLE of θ.

x	$f(x;1)$	$f(x;2)$	$f(x;3)$
0	1/3	1/4	0
1	1/3	1/4	0
2	0	1/4	1/4
3	1/6	1/4	1/2
4	1/6	0	1/4

If $x = 0$ is observed, $\hat{\theta} = 1$
If x=1, $\hat{\theta} = 1$.
if x=2, $\hat{\theta} = 2$ or 3.
if x=3, $\hat{\theta} = 3$.
if x=4, $\hat{\theta} = 3$.
Note here θ only takes three possible values. Given $X = x$ is observed, we find the θ value that gives largest $f(x;\theta)$. That is, find the θ value that maximizes the probability in each row.

exercise

2. Let $X_1, \cdots, X_n$ be a random sample from a gamma (α, β) distribution.
Find the MLE of β, assuming α known.
$f(x_1, \cdots, x_n; \beta) = \frac{\beta^{n\alpha}}{(\Gamma(\alpha))^n} (\prod x_i)^{\alpha-1} e^{-\beta \sum x_i}$,
Log $L(\beta) = n\alpha \log \beta - n \log \Gamma(\alpha) + (\alpha - 1) \log(\prod x_i) - \beta \sum x_i$,
$\frac{dLog(\beta)}{d\beta} = \frac{n\alpha}{\beta} - \sum x_i = 0$,
$\hat{\beta} = \frac{\alpha}{\bar{X}}$.

3. Given a random sample of size n from a beta distribution with $\beta = 1$, use the method of maximum likelihood to find an estimator for α.

MLE:

$f(x_1, \cdots, x_n) = (\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)})^n (\prod x_i)^{\alpha-1}$.

$\log L(\alpha) = n \log \alpha - (\alpha - 1) \log \prod x_i$,

$\frac{d \log L(\alpha)}{d\alpha} = \frac{n}{\alpha} + \log \prod x_i = 0$,

$\hat{\alpha} = \frac{n}{-\sum logx_i}$.

Large sample properties of estimators

When we use T which is a function of data to estimate θ, we expect a good estimator T for θ should be close to θ,
or $E_\theta(T-\theta)^2$, the mean squared error (MSE) of T, is small.
(Here the expectation is taken with respect to the distribution of T given θ.)
The MSE can be decomposed as
$E(T-\theta)^2 = \sigma_T^2 + (E(T)-\theta)^2$, Note $(\sigma_T^2 = Var(T))$
$E(T)-\theta$ is called the **bias** of T. If The bias of T is 0, we say T is an **unbiased estimator** for θ.
so MSE equals the variance of T plus the squared bias of T.
example: $X_1, \cdots, X_n$ iid $\sim N(\theta, \sigma^2)$. Estimate θ assuming σ^2 known.
Compare two estimators: $T_1 = \bar{X}_n$, $T_2 = X_1$. Find their MSE.
$E(T_1-\theta)^2 = \frac{\sigma^2}{n}, E(T_2-\theta)^2 = \sigma^2$.
Both of them are unbiased, so MSE equals their variance.

Limit of a sequence

$\{a_n\} = \{1/2, 3/4, 7/8, 15/16, \cdots, 1 - 1/2^n, \cdots,\}$

$\lim_{n\to\infty} a_n = 1$ or $a_n \to 1$.

$\{b_n\} = \{1/2, 3/4, 1/4, 7/8, 1/8, 15/16, 1/16, \cdots,\}$

$\overline{\lim}_{n\to\infty} b_n = 1, \underline{\lim}_{n\to\infty} b_n = 0.$

The limit of a sequence exists when the lower and upper limits coincide.

MSE consistent estimator

An estimator T is said to be a (MSE) **consistent** estimator of θ if $\lim_{n\to\infty} E_\theta(T-\theta)^2 = 0$.
e.g., T_1 is (MSE) consistent as its MSE $\frac{\sigma^2}{n}$ tends to 0 when n goes to infinity.
MLE yields consistent estimators.

Small sample properties of estimators

Unbiased estimator:
An estimator T is said to be an **unbiased** estimator of θ if $E_\theta(T) = \theta$ for every θ. We call $E_\theta(T) - \theta$ the **bias** of T.

e.g., $E(S_n^2) = \sigma^2$. The sample variance is unbiased for population variance. Recall $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X}_n)^2$.
MLE of σ^2 is $\frac{1}{n}\sum(X_i - \bar{X}_n)^2$.
$E(\hat{\sigma}^2_{MLE}) = E\frac{n-1}{n}S_n^2 = \frac{n-1}{n}\sigma^2$,
the bias is $\frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2$, that is, the MLE for σ^2 has a negative bias.
If we have several unbiased estimators, which is better? Examine variance.

UMVUE

An estimator T is said to be **Uniform Minimum Variance Unbiased Estimator** (UMVUE) of θ if
1) T is unbiased of θ.
2). $Var(T) < \infty$.
3). For any other unbiased estimator $\tilde{T}$, of θ, $Var(T) < Var(\tilde{T})$.

How to find UMVUE and prove it?
We can use Cramer-Rao Lower Bound (CRLB) to find UMVUE.

exercise

1. Let $X_1, \cdots, X_n$ iid $\sim N(\theta, 1)$.
Consider two estimators of θ.
$\hat{\theta} = \bar{X}$ and $\hat{\theta} = \frac{1}{2}(X_1 + X_2)$.
Find the MSE of both estimators. Are they consistent?
2. $X_1, \cdots, X_n$ iid $\sim Poi(\lambda)$.
Find the MLE of λ and its MSE.
$\bar{X}$ is MLE.
$\bar{X}$ is unbiased for λ, and $Var(\bar{X}) = \frac{\lambda}{n}$.
3. $X_1, \cdots, X_n$ iid $\sim Bernoulli(p)$. Find the MLE of p and its MSE.
(Bernoulli(p) is binomial (1, p)).
$\bar{X}$ is the MLE.
$E(\bar{X}) = p$, and $Var(\bar{X}) = \frac{p(1-p)}{n}$.

Solutions

1. Note $E(\bar{X}) = \theta, E(\frac{1}{2}(X_1 + X_2)) = \theta$, so both estimators are unbiased for θ, and the MSE is variance of each estimator.
$Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{1}{n}$.
$Var(\frac{1}{2}X_1 + \frac{1}{2}X_2) = \frac{1}{4}Var(X_1) + \frac{1}{4}Var(X_2) = \frac{1}{4} * 1 + \frac{1}{4} * 1 = \frac{1}{2}$.
The estimator $\bar{X}$ is MSE consistent, $\frac{1}{2}(X_1 + X_2)$ is not as its MSE does not go to 0 when n goes to infinity.

2. The likelihood function is
$L(\lambda) = f(x_1; \lambda)f(x_2; \lambda) \cdots f(x_n; \lambda) = \prod \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = (\prod \frac{1}{x_i})e^{-n\lambda}\lambda^{\sum x_i}$,
and the loglikelihood function is
$\log L(\lambda) = \log(\prod \frac{1}{x_i}) - n\lambda + \sum x_i \log(\lambda)$,
Let $\frac{d \log L(\lambda)}{d\lambda} = -n + \frac{\sum x_i}{\lambda} = 0$,
we can solve $\hat{\lambda} = \frac{\sum X_i}{n} = \bar{X}_n$.
i.e., the MLE for λ is the sample mean X_n.
$E(\bar{X}_n) = \lambda$, so MSE$=Var(\bar{X}_n) = \frac{\lambda}{n}$ hence $\bar{X}_n$ is MSE consistent.

Solutions

3. The likelihood function is
$L(p) = f(x_1; p) f(x_2; p) \cdots f(x_n; p) = \prod p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}$,
The loglikelihood function is
$\log L(p) = \sum x_i \log(p) + (n - \sum x_i) \log(1-p)$,
Let $\frac{d \log L(p)}{dp} = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = 0$,
we can solve $\hat{p} = \bar{X}_n$
$\bar{X}_n$ is unbiased for p and its MSE $= Var(\bar{X}_n) = \frac{p(1-p)}{n}$,
so $\bar{X}_n$ is MSE consistent.
Note if $X \sim Bernoulli(p)$, then $E(X) = p$, $Var(X) = p(1-p)$.