It is a truth very certain that when it is not in our power to determine what is true we ought to follow what is most probable. –René Descartes

Problems 31-40 on Likelihood (STAT2802 Statistical Models Tutorial notes for the week of 15-OCT-2012)

We denote by $f_X(x;\theta)$ the density, carrying a parameter θ, on the sample space of the random variable X. With such notation are the assumptions about variability of x and fixation of θ. But these assumptions are imposed externally by the statistician rather than arisen self-evidently from the mathematical expression. The mathematical expression of $f_X(x;\theta)$ is really just $f(x,\theta)$—a function taking two arguments (assuming univariate-uniparametric case)—with both arguments variable. Thus, a dual interpretation of the mathematical expression is to constrain the first argument, x, at a single fixed point and let the second argument, θ, vary.This dual interpretation describes a face of reality: X is fixed by observation at x and the parameter is an unknown *variable* in the parameter space. To describe this mathematical consideration with more intuition, we name it *the likelihood of the parameter given by the data* and denote by $L(\theta;x)$ or simply $L(\theta)$. The principle for formal manipulation of $L(\theta)$ is that to still view it as the bivariate function $f(x,\theta)$ and remember that it is ≥0 and integrates to 1 with respect to x, but it does not integrate to 1 with respect to θ.

To describe $L(\theta;x)$ as the likelihood of θ given x is to say that the likelihood function will evaluate on any specified θ_0 to a numerical value representing θ_0's likelihood. Thus, if $L(\theta_1;x) > L(\theta_2;x)$, then we say that θ_1 has a higher likelihood than θ_2 given the same set of observed data. If we push along this line of intuition, we will reach a maximum likelihood estimator $\hat{\theta}_m(x) = \arg\max_{\theta\in\Theta} L(\theta;x)$.

When sample size is very large, the Maximum Likelihood Estimator is an estimator that enjoys many qualities (consistency, normality, efficiency, unbiasedness, sufficiency) of an ideal estimator. It is the most theoretically important, general method of constructing an estimator.

Problems 31-40

31. From a large lake containing an unknown number N of fish, a random sample of M fish is taken. The fish caught are marked with red spots and released into the lake. After some time, another random sample of n fish is drawn and it is observed that k of them are spotted. Assuming $N \gg M$ and n. Show that $\Pr(k;N,M,n)$ the probability that the second sample contains exactly k spotted fish, is given by the hyper-geometric distribution

$$\Pr(k;N,M,n) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

By considering the ratio $\frac{L(N|M,n;k)}{L(N-1|M,n;k)}$, deduce that the maximum likelihood estimate of N is the largest integer short of $\frac{nM}{k}$.

32. For the log-Normal distribution defined by the probability density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\ln x-\mu)^2}\mathbb{I}(X \geq 0),$$

show that the maximum-likelihood estimators of μ and σ^2 are

$$\hat{\mu}_{\text{mle}} = g \text{ and } \hat{\sigma}^2_{\text{mle}} = \frac{1}{n}\sum_{i=1}^{n}(\ln X_i - g)^2,$$

where $g = \frac{1}{n}\sum_{i=1}^{n} \ln X_i$ is the logarithm of the geometric mean of the size-n random sample.

33. A continuous random variable X defined in the range $[0, +\infty)$ has a density function proportional to $xe^{-x/\theta}$, $\theta > 0$. Find the mean and variance of X. If a random sample of size n is drawn from this population, obtain the maximum-likelihood estimate of the parameter θ and calculate the variance of the estimate.

34. In an animal-breeding experiment four distinct kinds of progeny were observed with the frequencies n_1, n_2, n_3 and n_4 ($\sum n_i \equiv N$). The corresponding expected proportions on a biological hypothesis are $\frac{1}{4}(2+p), \frac{1}{4}(1-p), \frac{1}{4}(1-p), \frac{1}{4}p$, where p is an unknown parameter. Obtain $\hat{p}_{\text{mle}}$ for p and verify that its large-sample variance is $\frac{2p(1-p)(2+p)}{N(1+2p)}$.

35. A Γ variable X has the probability density function

$$f_X(x) = \frac{1}{a\Gamma(p)} e^{-\frac{x}{a}} \left(\frac{x}{a}\right)^{p-1}, \qquad \text{for } X \geq 0.$$

Given n independent observations $x_1, x_2, \ldots, x_n$ of X, prove that the expectations of the sample arithmetic and geometric means are

$$ap \text{ and } a\left[\frac{\Gamma\left(p+\frac{1}{n}\right)}{\Gamma(p)}\right]^n \quad \text{respectively.}$$

Hence deduce that the ratio of the population arithmetic and geometric mean (defined as $\lim_{n\to\infty} \mathbb{E}\left(\prod_{i=1}^{n} X_i^{\frac{1}{n}}\right)$ is

$$\theta := pe^{-\phi(p)}, \qquad \text{where } \phi(p) \equiv \frac{d}{dp}[\ln\Gamma(p)].$$

Also, show that $\hat{\theta}_{mle}$, the maximum likelihood estimator for θ, is the ratio of the sample arithmetic and geometric means. *Hint: You may use the fact that θ is a strictly decreasing function of p when $p \in (0,1)$.*

36. Find the maximum likelihood estimate of the parameter p of a Bernoulli(p) population using a random sample of size n and derive the estimator's variance.

37. Find the maximum likelihood estimate of the parameter λ of a Poisson(λ) population using a random sample of size n and derive the estimator's variance.

38. Find the maximum likelihood estimate of the parameter p of a Geometric(p) population using a random sample of size 1 and derive the estimator's bias.

39. If $\hat{\theta}$ is the maximum likelihood estimate of a parameter θ and $\varphi(\theta)$ is a strictly monotonically increasing function of θ, show that $\varphi(\hat{\theta})$ is the maximum likelihood estimate of $\varphi(\theta)$. Then find the maximum likelihood estimate of the 4th central moment of the normal distribution which is equal to $3\sigma^4$.

40. Suppose that a box contains ten balls, and let p be the proportion of balls that are red. Two balls are drawn with replacement. Find the probability function $p_X(x)$ of the r.v. $X = \textit{number of red balls drawn}$. For each x that X can assume determine the value of p that maximizes $f(x)$.