From a theoretical point of view, the most important general method of estimation so far known is the method of maximum likelihood. –Harald Cramér

Solutions to 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. Show that $\Pr(k; N, M, n)$ the probability that the second sample contains exactly k spotted fish, is given by the hyper-geometric p.m.f.

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

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 $\frac{nM}{k}$.

Solution. From the description of the experiment, we know that the parameters M and n are under control by the statistician, therefore they are known constants of the model and the only unknown parameter of the model is N, the total number of fish in the lake. Also from the description of the experiment, the true sample size is 1, i.e., the hyper-geometric random variable is only observed *once*, as the integer k. The likelihood of the parameter N given by the single datum k is in the same analytic form of the univariate hyper-geometric p.m.f. (after fixing parameters M and n to constants as discussed):

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

To maximize this integer domain function (as N has to be an integer), we nevertheless could consider the incremental ratio

$$\frac{L(N|M, n; k)}{L(N-1|M, n; k)} = \frac{\binom{N-M}{n-k}}{\binom{N-M-1}{n-k}} \frac{\binom{N-1}{n}}{\binom{N}{n}} = \frac{N-M}{N} \frac{N-n}{N-M-n+k}.$$

When this ratio, as a function of N, is nearest to unity, that is, when

$$(N-M)(N-n) = N(N-M-n+k) \Leftrightarrow N = \frac{nM}{k},$$

the likelihood is at maximum. For integer N, this means the last incremental ratio (>1) before decrementing is $\left\lfloor \frac{nM}{k} \right\rfloor$ 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.

Solution. The likelihood of the parameters given by a size-n random sample has the same analytic form as the product of n univariate densities—it is the interpretation of the analytic form that distinguishes the likelihood from the joint density. Likelihood (all $X_i \geq 0$): $L(\theta; \{X_i\}_1^n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln X_i-\mu)^2}$. Log-likelihood: $\ell(\theta; \{X_i\}_1^n) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\ln X_i - \mu)^2$. Likelihood Equations:

$$\begin{cases} 0 = \dfrac{\partial \ell(\theta)}{\partial \mu} = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(\ln X_i - \mu) \Rightarrow \mu = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\ln X_i = g \\ 0 = \dfrac{\partial \ell(\theta)}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2(\sigma^2)^2}\displaystyle\sum_{i=1}^{n}(\ln X_i - \mu)^2 \Rightarrow \sigma^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(\ln X_i - \mu)^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(\ln X_i - g)^2 \end{cases}$$

The solutions to these two likelihood equations represent the analytic form of the MLEs of the respective parameters, note that g is a statistic and does not depend on the parameters when the sample is given.

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.

Solution. The normalizing constant is $\int_0^\infty xe^{-\frac{x}{\theta}}dx = \theta^2\int_0^\infty ye^{-y}dy = \theta^2\Gamma(2) = \theta^2$. Therefore the full specification of the density of $X(\geq 0)$ is $X \sim f_X(x) = \frac{1}{\theta^2}xe^{-\frac{x}{\theta}}\mathbb{I}(x \geq 0, \theta > 0)$. Hence $\mathbb{E}(X) = \frac{1}{\theta^2}\int_0^\infty x^2e^{-\frac{x}{\theta}}dx = \theta\int_0^\infty y^2e^{-y}dy = \theta\Gamma(3) = 2\theta$ and $\mathbb{E}(X^2) = \frac{1}{\theta^2}\int_0^\infty x^3e^{-\frac{x}{\theta}}dx = \theta^2\int_0^\infty y^3e^{-y}dy = \theta^2\Gamma(4) = 6\theta^2$ and $\mathbb{V}(X) = 2\theta^2$. The likelihood of the size-n random sample is in the same analytic form as the product of the n univariate density (all $X_i \geq 0$): $L(\theta; \{X_i\}_1^n) =$

$\left(\frac{1}{\theta^2}\right)^n (\prod_{i=1}^n X_i) e^{-\frac{1}{\theta}\sum_{i=1}^n X_i}$ and log-likelihood is: $\ell(\theta) = -2n\ln\theta + \sum_{i=1}^n \ln X_i - \frac{1}{\theta}\sum_{i=1}^n X_i$. Likelihood equation: $0 = \frac{\partial \ell}{\partial \theta} = -\frac{2n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^n X_i \Rightarrow \theta = \frac{1}{2n}\sum_{i=1}^n X_i$ which represents the analytic form of $\hat{\theta}_{\text{mle}}$. Note that it is an unbiased estimator for the parameter θ. Variance of the MLE is $\mathbb{V}(\hat{\theta}_{mle}) = \mathbb{V}\left(\frac{1}{2n}\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \frac{1}{4n^2}\mathbb{V}(X_i) = \frac{2\theta^2}{4n} = \frac{\theta^2}{2n}$. Note that the variance tends to 0 as $n \to \infty$, with unbiasedness, this means that it is also a consistent estimator for the parameter θ.

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 $\frac{2p(1-p)(2+p)}{N(1+2p)}$.

Solution. The description of the experiment clearly points to a (reparametrized) quadrinomial distribution: $p_X(n_1, n_2, n_3, n_4(\sum n_i = n); N, p) = \binom{N}{n_1, n_2, n_3, n_4}\left[\frac{1}{4}(2+p)\right]^{n_1}\left[\frac{1}{4}(1-p)\right]^{n_2}\left[\frac{1}{4}(1-p)\right]^{n_3}\left[\frac{1}{4}p\right]^{n_4}$. The true sample size with respect to the multinomial perspective is 1 and the likelihood is in the same analytic form as the p.m.f.. Since N is observed by the statistician, so it is no longer a parameter but a constant. The only true unknown parameter is p. Log-likelihood:

$$\ell\big(p \big| N; n_1, n_2, n_3, n_4(\textstyle\sum n_i = n)\big) = \ln c + n_1 \ln\frac{2+p}{4} + n_2 \ln\frac{1-p}{4} + n_3 \ln\frac{1-p}{4} + n_4 \ln\frac{p}{4}$$

Likelihood equation:

$$0 = \frac{\partial \ell}{\partial p} = \frac{n_1}{2+p} - \frac{n_2+n_3}{1-p} + \frac{n_4}{p} \Rightarrow \hat{p}_{\text{mle}} = \left(-1 + \frac{3n_1+n_4}{2N}\right) + \sqrt{\frac{1}{4}\left(-1+\frac{3n_1+n_4}{2N}\right)^2 + \frac{2n_4}{N}}$$

The large-sample variance of MLE is equal to $1/i(\theta)$ where $i(\theta) = -\mathbb{E}\left(\frac{\partial^2 \ell(p)}{\partial p^2}\right) = \frac{1}{4}\frac{N}{2+p} + \frac{1}{2}\frac{N}{(1-p)} + \frac{1}{4}\frac{N}{p} = \frac{N(1+2p)}{p(2+p)(1-p)}$ because $\mathbb{E}(n_1) = N \cdot \frac{1}{4}(2+p)$ and so on.

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 $x_1, x_2, \dots, x_n$ 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 $\hat{\theta}_{\text{mle}}$, the maximum likelihood estimator for θ, is the ratio of the sample arithmetic and geometric means.

Solution. $\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) = \mathbb{E}(X) = \int_0^\infty \frac{1}{a\Gamma(p)} x e^{-\frac{x}{a}} \left(\frac{x}{a}\right)^{p-1} dx = \frac{a}{\Gamma(p)} \int_0^\infty e^{-y} y^p dy = \frac{a\Gamma(p+1)}{\Gamma(p)} = ap$

$$\mathbb{E}\left(\prod_{i=1}^n X_i^{\frac{1}{n}}\right) = \prod_{i=1}^n \left(\mathbb{E}X_i^{\frac{1}{n}}\right) = \left(\mathbb{E}X_1^{\frac{1}{n}}\right)^n = \left[\int_0^\infty \frac{1}{a\Gamma(p)} x^{\frac{1}{n}} e^{-\frac{x}{a}} \left(\frac{x}{a}\right)^{p-1} dx\right]^n = \left[\frac{a^{\frac{1}{n}}}{\Gamma(p)} \int_0^\infty e^{-y} y^{p-1+\frac{1}{n}} dy\right]^n = a\left[\frac{\Gamma\left(p+\frac{1}{n}\right)}{\Gamma(p)}\right]^n.$$

The population geometric mean is $\lim_{n\to\infty} \mathbb{E}\left(\prod_{i=1}^n X_i^{\frac{1}{n}}\right) = \lim_{n\to\infty} a\left[\frac{\Gamma\left(p+\frac{1}{n}\right)}{\Gamma(p)}\right]^n$. As $n \to \infty$, $n\ln\frac{\Gamma\left(p+\frac{1}{n}\right)}{\Gamma(p)} = n\ln\frac{\Gamma(p)+\frac{1}{n}\Gamma'(p)}{\Gamma(p)} = n\ln\left(1+\frac{1}{n}\frac{\Gamma'(p)}{\Gamma(p)}\right) = n\frac{1}{n}\frac{\Gamma'(p)}{\Gamma(p)} = \frac{d}{dp}\ln\Gamma(p) = \phi(p)$.

Since the relationship between θ and p is one-to-one, $\hat{\theta}_{mle} = \theta(\hat{p}_{mle})$.

Log-Likelihood of p:

$$\ell(p) = -n\ln a - n\ln\Gamma(p) - \sum_{i=1}^n \frac{X_i}{a} + (p-1)\sum_{i=1}^n \ln\frac{X_i}{a} = -np\ln a - n\ln\Gamma(p) - \frac{n}{a}\bar{X}_n + n(p-1)\ln G_n$$

where G_n is the sample geometric mean. Likelihood equation for p:

$$0 = \frac{\partial\ell}{\partial p} = -n\ln a - n\phi(p) + n\ln G_n \Rightarrow \phi(p) = \ln\frac{G_n}{a} \Rightarrow pe^{-\phi(p)} = \frac{ap}{G_n}.$$

Therefore $\hat{\theta}_{mle} = \bar{X}_n / G_n$.

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.

Solution. Likelihood: $L(p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i}$. Log-Likelihood: $\ell(p) = \ln p \sum_{i=1}^{n} X_i + \ln(1-p) \sum_{i=1}^{n}(1-X_i)$. Likelihood Equation for unknown p: $0 = \frac{\partial \ell}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} X_i - \frac{1}{1-p}\sum_{i=1}^{n}(1-X_i) \Rightarrow p = \bar{X}_n$. Hence $\hat{p}_{\text{mle}} = \bar{X}_n$. (unbiased for p). $\mathbb{V}(\hat{p}_{\text{mle}}) = \frac{1}{n}\mathbb{V}(X_1) = \frac{1}{n}p(1-p)$. (with unbiasedness, the estimator is consistent for p).

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.

Solution. Likelihood: $L(\lambda) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{X_i}}{X_i!}$. Log-Likelihood: $\ell(\lambda) = \sum_{i=1}^{n}(-\lambda + X_i\ln\lambda - \ln(X_i!))$. Likelihood Equation for unknown λ: $0 = \frac{\partial \ell}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} X_i \Rightarrow \lambda = \bar{X}_n$. Hence $\hat{\lambda}_{\text{mle}} = \bar{X}_n$. (unbiased for λ). $\mathbb{V}(\hat{\lambda}_{mle}) = \frac{1}{n}\mathbb{V}(X_1) = \frac{1}{n}\lambda$. (with unbaisedness, the estimator is consistent for λ).

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

Solution. Population distribution: $p_X(X=k) = (1-p)^k p$, for $k = 0,1,2,\ldots$. Likelihood: $L(p) = \prod_{i=1}^{n}(1-p)^{X_i}p$. Log-likelihood: $\ell(p) = \sum_{i=1}^{n}(X_i\ln(1-p) + \ln p)$. Likelihood equation for p: $0 = \frac{\partial \ell}{\partial p} = \sum_{i=1}^{n}(-\frac{X_i}{1-p} + \frac{1}{p}) \Rightarrow p = \frac{1}{1+\bar{X}_n}$. Hence $\hat{p}_{\text{mle}} = \frac{1}{1+\bar{X}_n}$ and for $n = 1$, it is $\hat{p}_{\text{mle}} = \frac{1}{1+X_1}$. $\mathbb{E}(\hat{p}_{\text{mle}}) = \sum_{k=0}^{\infty}\frac{q^k p}{1+k} = \frac{p}{q}\sum_{k=0}^{\infty}\frac{q^{1+k}}{1+k} = \frac{p}{q}\sum_{k=0}^{\infty}\int_0^q x^k dx = \frac{p}{q}\int_0^q\left(\sum_{k=0}^{\infty}x^k\right)dx = \frac{p}{q}\int_0^q \frac{dx}{1-x} = -\frac{p}{q}\ln(1-q) = \frac{p\ln p}{p-1}$. Hence $b(\hat{p}_{\text{mle}}) = \frac{p\ln p}{p-1} - p$.

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 $3\sigma^4$.

Solution. (Special case of the invariance principle.) Since φ is a one-to-one function, it has an inverse function φ^{-1}. Write $\theta = \varphi^{-1}(\tau)$. Then

$$L(\theta) = L \circ \varphi^{-1}(\tau) = L \circ \varphi^{-1} \circ \varphi(\theta) \leq L \circ \varphi^{-1} \circ \varphi(\hat{\theta})$$

Note that $\varphi(\hat{\theta})$ is on the range of τ and $L \circ \varphi^{-1}$ is the likelihood function in terms of τ. This means $\hat{\tau} = \varphi(\hat{\theta})$.

The MLE of $3\sigma^4$ is just $3\left(\widehat{\sigma^2_{mle}}\right)^2$ where $\widehat{\sigma^2_{mle}} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$. This is because the relationship between $\sigma^4 \sim \sigma^2$ is strictly monotonically increasing on $(0, +\infty)$. Hence $\mathrm{MLE}(3\sigma^4) = 3\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right)^2$.

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

Solution.

$\boldsymbol{x}$	$\mathbf{0}$	$\mathbf{1}$	$\mathbf{2}$
$\boldsymbol{p_X(x)}$	$(1-p)^2$	$2p(1-p)$	p^2
$\hat{\boldsymbol{p}}$	0	$1/2$	1