8. Newcomb, S. (1903). *Reminiscences of an Astronomer*. Harper, London and New York. (Reprinted in its entirety in Stigler [14].)
9. Norberg, Arthur L. (1978). *Isis*, 69, 209–225. (Discusses Newcomb's astronomical career up to 1870).
10. Raimi, R. (1976). *Amer. Math. Monthly*, **83**, 521–538. (A review article on the leading digit distribution.)
11. Rubin, E. (1967). *Amer. Statist.*, October, 45–48. (Discusses Newcomb's work on the sex ratio at birth.)
12. Stigler, S. M. (1973). *J. Amer. Statist. Ass.*, **68**, 872–879. (Discusses Newcomb's work on robust estimation. Reprinted in Stigler [14].)
13. Stigler, S. M. (1977). *Ann. Statist.*, **6**, 239–265. (Quotes Newcomb on sufficiency and discusses his place in early American work. Reprinted in Stigler [14].)
14. Stigler, S. M., ed. (1980). *American Contributions to Mathematical Statistics in the Nineteenth Century*, two volumes. Arno Press, New York. (Includes photographic reprints of several of Newcomb's works in statistics as well as the whole of his autobiography.)

See also ASTRONOMY, STATISTICS IN.

STEPHEN M. STIGLER

NEWSBOY INVENTORY PROBLEM

A definition and classical formulation of the newsboy inventory model as a profit maximization problem is provided. The structure of the optimal stocking policy is given. The alternative minimax formulation for the distribution-free newsboy model is also presented. Demand estimation in the presence of fully observable and censored sales is discussed from the frequentist and Bayesian perspectives. Explicit formulas are provided for Bayesian updating of a comprehensive set of demand functions.

The newsboy problem refers to the determination of the optimal ordering (stocking) quantity based on the trade-off between excess inventory and shortage costs for products with useful lives of only one planning period. It is also called *the Christmas tree problem*, *single period inventory problem*, and *the newsvendor problem*. It is directly applicable when a product perishes quickly such as fresh produce, certain style goods, and newspapers (hence, the name). The newsboy model is the building block for stochastic dynamic inventory problems of longer horizons where, at the end of one period, another period begins with the leftover inventory from the previous period as the initial inventory in the current one. Moreover, with the appropriate choice of excess and shortage costs incurred at the end of a single period, it also provides a good myopic approximation for an infinite horizon inventory problem with positive delivery lead times and lost sales [15]. For certain inventory systems, the myopic policy has been established to be the optimal policy [20]. Hence, the structural properties of the newsboy problem deserve attention aside from its immediate applicability in single-period settings.

The earliest analysis of the newsboy problem is by Arrow [2]. The planning horizon consists of a single selling time period in which there is only one purchasing (stocking) opportunity at the beginning of the period with instantaneous delivery of purchased items. The demand for the product during the selling period is a random variable X which has known cumulative distribution function (cdf) $F(x)$ and probability density function (pdf) $f(x)$, with known parameters. For convenience, assume that X is purely continuous. Similar results hold when X is discrete or of a mixed nature. Each unit purchased costs c, each unit sold brings in a revenue of r, each unit disposed as salvage gives a revenue of r', and there is a penalty cost of p per unit of unsatisfied demand. Associated with the order, there is a fixed ordering cost K. All cost parameters are nonnegative. Suppose the inventory on hand at the start of the period *before* ordering is $I_0 \geqslant 0$. The decision variable, S, is the inventory level *after* ordering. Hence, S is often called *the order-up-to-point* and satisfies $S \geqslant I_0$. The newsboy problem can be formulated either as a cost minimization or a profit maximization problem; the two formulations give equivalent results. Herein, a profit maximization is presented. Let $G(S, I_0)$ denote the expected total profit with initial inventory I_0 and S units on hand after ordering. The optimization problem is stated formally as

$$\max_S G(S, I_0) = r \iint_0^\infty \min(S, \xi) f(\xi) d\xi + r\prime \int_0^S (S - \xi) f(\xi) d\xi - c(S - I_0) - K\delta[S - I_0] - p \int_S^\infty (\xi - S) f(\xi) d\xi \quad (1)$$

where $\delta[S - I_0] = 1$ if $S > I_0$, and zero otherwise. It is easy to show that $G(S, I_0)$ is convex in S so that the optimizing value of S, say S^*, occurs where

$$F(S^*) = \frac{r - c + p}{r - r\prime + p} = \gamma.$$

Hence, the maximum stocking level at the beginning of the period cannot exceed S^*. In addition, when there are positive fixed ordering costs, there will be an optimal reorder point, s^*, such that an order will be placed if and only if the initial inventory is below it. The optimal reorder point is that initial inventory level that results in the same expected profit with no additional purchasing as that obtained with S^* units on hand. Thus, the optimal solution: if $I_0 < s^*$ order $S^* - I_0$, otherwise do not order. Typically, it is assumed that the fixed ordering cost is negligible so that $s^* = S^*$. In this case, the optimal solution reduces to a single critical number policy. Since $F(S^*)$ is the probability that demand does not exceed S^*, the optimal solution occurs where this probability is equal to the critical ratio, γ, which is often expressed in the form $c_u/(c_u + c_o)$, where c_u is the underage cost (i.e., $r - c + p$), and c_o is the overage cost (i.e., $c - r'$). The critical ratio is also called *the desired service level*. When X is a discrete random variable, the optimal order-up-to point is the smallest integer such that the desired service level is satisfied.

The most commonly used distributions are normal for continuous demand, and Poisson and negative binomial for discrete demand; the last providing a better fit empirically for retail data [1]. Scarf [17] addressed the newsboy problem where only the first two moments of the demand distribution are known (μ and σ^2) without any further assumptions about the form of the distribution. This version of the problem is called *the distribution-free newsboy problem* under the minimax criterion. Formally, it is stated as

$$\max_S \min_{f(\cdot)} \left[r \iint_0^\infty \min(S, \xi) f(\xi) d\xi + r\prime \iint_0^S (S - \xi) f(\xi) d\xi - cS \right]$$

where the minimization over the functions $f(\cdot)$ is subject to $\iint \xi f(\xi) d\xi = \mu$ and $\iint (\xi - \mu)^2 f(\xi) d\xi = \sigma^2$. The worst distribution of demand is found to have positive mass at only two points, say a and b, with its mean and standard deviation equal to the given values, μ and σ. The optimal stocking quantity is then given by

$$S^* = \begin{cases} 0 \left(\frac{c-r'}{r-r'}\right)\left(1 + \frac{\sigma^2}{\mu^2}\right) > 1 \\ \mu + \sigma h(\frac{c-r'}{r-r'}) \left(\frac{c-r'}{r-r'}\right)\left(1 + \frac{\sigma^2}{\mu^2}\right) < 1 \end{cases},$$

where $h(x) = 1/2 - x[x(1-x)]^{1/2}$. Furthermore, if this policy is used, the minimum expected profit is $\max\{[(r-c)\mu - \sigma[c(1-c)]^{1/2}], 0\}$ for *all* distributions with mean μ and standard deviation σ. Other works on the distribution-free newsboy problem are References 6 and 14. Lowe, Schwarz, and McGavin [13] investigate the newsboy problem when there is uncertainty about the underage and overage costs expressed as an interval for each. For the classical setting, the optimal single-number policy determined through the critical ratio still holds if the costs are replaced by their expected values. Under the minimax criterion, there are two optima determined through the critical ratio computed by either the lowest underage and highest overage costs or the highest underage and lowest overage costs.

Estimation of the distribution of demand during the selling period is of interest. For commodity-type perishable products, a historically observed histogram of total demands can be used as a direct estimate of the probability distribution of demand in a future period. Similarly, past data can be used only to estimate the parameters of the demand distribution that is assumed of a certain form. In the presence of fully observed demands, obtaining the maximum likelihood estimator (m.l.e.) is quite straightforward.

When there are unobservable demands (unrecorded lost sales), the estimation procedure consists of the EM algorithm, and usually gets computationally cumbersome. For Poisson demand with unknown mean λ, Conrad [5] derives a maximum likelihood estimator of the mean. For the negative binomial demand, Agrawal and Smith [1] develop a procedure based on the uncensored MLEs provided by Johnson and Kotz [11]. For normal demand, m.l.e. estimators [8] and the approximate BLUE estimators [7] for the mean and standard deviation are available, albeit computationally intensive. For normal demand, Nahmias [16] proposes much simpler estimators, which may be of great use to practitioners. Let $(x_1, x_2, \cdots, x_n)$ be a random sample from a normal population with unknown mean μ and unknown standard deviation σ. Assume that S is a known constant and only the values of the sample for which $x_i < S$ are observed. Define ρ as the order of the maximum observed value. For notational convenience, assume that the sample values are ordered, so that $x_1, x_2, \ldots, x_\rho$ are observed sample values. Then, the simplified estimators for the standard deviation and the mean, respectively ($\tilde{\sigma}$ and $\tilde{\mu}$), are

$$\tilde{\sigma} = \frac{v_\rho^2}{1 - z\phi(z)/u - [\phi(z)/u]^2},$$

$$\tilde{\mu} = \overline{x}_\rho + \frac{\tilde{\sigma}\phi(z)}{u},$$

where

$$z = \Phi^{-1}(u), u = \rho/n, \overline{x}_\rho = (1/\rho)\sum_{i=1}^{\rho} x_i,$$

$$v_\rho^2 = \frac{1}{\rho - 1}\sum_{i=1}^{\rho}(x_i - \overline{x}_\rho)^2,$$

and ϕ and Φ are the standard normal density and cumulative distribution functions, respectively. Both estimators are only functions of three sample statistics and are easily computed. For large n and $S \geqslant \mu$, the proposed estimators and the MLEs have very similar characteristics.

For some products, such as style goods or high-end electronic goods, historical data may be lacking or, at best, not likely to be representative of future conditions. Then, Bayesian procedures are applicable. Although the Bayesian analysis is more appropriate for multiple-period dynamic inventory problems, it may be used in the newsboy setting as well, if it is possible to observe a number of demands before a final-stocking decision is made. In the Bayesian approach, one or more parameters of the probability distribution of demand are assumed unknown with precision but prior knowledge is encoded in the form of probability distributions over possible values of these parameters. Details of how such priors may be generated by eliciting information from experts are provided in Reference 19. Let the parameter θ be a random variable (or vector) with *prior* pdf $\pi(\theta)$, and $f(x|\theta)$ denote the conditional density of X given θ. After the demand data $\mathbf{D}$ (typically, of previously recorded n demands $\xi_1, \xi_2, \cdots, \xi_n$) is observed, the density of θ is updated to obtain the *posterior* pdf as $g(\theta|\mathbf{D})$, from which the posterior density of demand is found as

$$f(x|\mathbf{D}) = \int f(x|\theta)g(\theta|\mathbf{D})d\theta.$$

The above density is then used in the optimization. For certain types of likelihood functions, the prior density can be selected from a particular family---the conjugate family---so that the updating operation results in the posterior density being another member of the same family. Although the computationally efficient techniques have provided great relief, the choice of the prior within the conjugate family is still a necessity for analytically tractable solutions; and, hence, in the sequel, only the results with conjugacy property are summarized.

Scarf [18] provides the first inventory theoretic analysis of Bayesian demand estimation. The objective is to minimize expected total costs over a finite horizon where the fixed ordering costs in a period are ignored. The demand during a period is assumed to have the density within the exponential class of the form $f(x|\omega) = \beta(\omega)e^{-\omega x}r(x)$, where the unknown parameter ω has the prior density denoted by $\pi(\omega)$. The demand data $\mathbf{D}$ may be summarized in the sufficient statistic $y = \sum_{i=1}^{n} \xi_i/n$. Thus, the posterior demand density is given by

$$f(x|y) = r(x) \int_0^\infty \beta^{n+1}(\omega) e^{-\omega x} e^{-n\omega y} \pi(\omega) d\omega \times \left[\int_0^\infty \beta^n(\omega) e^{-\omega y} \pi(\omega) d\omega \right]^{-1}.$$

The optimal stocking policy is of the order-up-to type, and the optimal stocking levels are found to be increasing in the sufficient statistic y. Iglehart [10] extends the analysis to include the case when $f(x|\omega)$ is a member of the range family where

$$f(x|\omega) = q(x)r(\omega)\psi(x, \omega)$$

with $q(x) = 0$ for $x < 0$, and $\psi(x, \omega) = 1$ if $x \leqslant \omega$ and zero otherwise. This family includes all continuous random variables whose support is $(0, \omega)$, the simplest example of which is the uniform random variable. The sufficient statistic for this family is $y = \max_{1 \leqslant i \leqslant n}\{\xi_i\}$. A comprehensive study on Bayesian solution of dynamic inventory problems is by Azoury [3]. It is shown that, for certain classes of demand densities, there exists a function $q(y)$ of the sufficient statistic y, which is a scale parameter on the demand density such that the expected cost function over a finite horizon is scalable in the observed demand $\mathbf{D}$. Thus, it suffices to solve a "standardized" optimization problem and the optimal stocking quantity is $S^* = \gamma q(y)$, where γ is the optimal solution for the standardized problem. The scalability property is extremely useful in solving the dynamic inventory problems because it enables one to reduce greatly the state of the system that arises from the possible demand observations over the horizon. The basic results for the scalable demand densities are given below.

When the demand has a uniform distribution with unknown parameter ω and the prior on ω is a Pareto distribution with density function $\pi(\omega) = aR^a/\omega^{a+1}$, where $\omega > R$ and a and R are positive, the posterior demand distribution is given by

$$f(x|y) = \frac{(a+n)[q(y)]^{a+n}}{(a+n+1)[\max\{x, q(y)\}]^{a+n+1}},$$

where $y = \max_{1 \leqslant i \leqslant n}\{\xi_i\}$, and $q(y) = \max(y, R)$.

When the demand distribution belongs to the Weibull family with $f(x|\omega) = \omega k x^{k-1} \exp(\omega x^k)$, where k is the known shape parameter and ω is the unknown scale parameter with gamma prior with known parameters a and b, the sufficient statistic is $y = \sum_{i=1}^n \xi_i^k$, $q(y) = (b+y)^{1/k}$, and

$$f(x|y) = \frac{k(a+n)(b+y)^{a+n} x^{k-1}}{(b+y+x^k)^{a+n+1}}.$$

Finally, when the demand in period i is given by $x_i = k_i Z$ and Z has a gamma distribution with the known shape parameter λ and the unknown scale parameter ω with gamma prior defined as above, the sufficient statistic is $y = \sum_{i=1}^n \xi_i/k_i$, $q(y) = k_n(b+y)$, and the posterior demand density for the next period is

$$f(x|y) = \frac{\Gamma(a+(n+1)\lambda)(b+y)^{a+n\lambda}(x/k_{n+1})^{\lambda-1}}{k_{n+1}\Gamma(\lambda)\Gamma(a+n\lambda)(b+y+x/k_{n+1})^{a+(n+1)\lambda}}$$

where $\Gamma(\cdot)$ is the gamma function.

When the demand has an exponential distribution with unknown parameter λ and the prior on λ is noninformative [9], the posterior demand distribution is a Pareto distribution:

$$f(x|y) = \frac{(n+1)y^{(n+1)}}{(x+y)^{(n+2)}}.$$

In this particular case, the optimal stocking quantity has a simple form: $S^* = y(\gamma^{-1/(n+1)} - 1)$, where γ is the critical ratio.

When the demand is Poisson with unknown rate λ and the prior on λ is noninformative [9], the posterior distribution for λ is gamma:

$$g(\lambda|y) = \frac{n^{(y+1)}\lambda^y e^{-ny}}{y!}$$

Then, the posterior demand distribution belongs to the negative binomial family with parameters $(y+1)$ and $n/(n+1)$, where the probabilities are computed recursively, using

$$f(0|y) = \left(\frac{n}{n+1}\right)^{y+1}$$

and for $x = 1, 2, \ldots$

$$f(x|y) = \frac{y + x}{x(n+1)} f(x-1|y)$$

When the demand has a binomial distribution with known N but unknown p and the prior on p is uniform [9], the posterior distribution of p is the beta density:

$$g(p|y) = \frac{(nN+1)!p^y(1-p)^{nN-y}}{y!(nN-y)!}.$$

Then, the posterior demand distribution belongs to the hypergeometric family and is obtained recursively as

$$f(0|y) = \frac{(nN+1)!((n+1)N-y)!}{((n+1)N+1)!(nN-y)!}$$

and for $x = 1, 2, \ldots, N$

$$f(x|y) = \frac{(N-x+1)(x+y)}{x((n+1)N-y-x+1)} f(x-1|y).$$

Berk, Gürler, and Levine [4] consider the case when the demand has gamma distribution with the known shape parameter α and the unknown scale parameter β with gamma prior with initial shape and scale parameters ρ_0 and τ_0. With the cumulative demand y over the last n observations, the posterior density of demand is Gamma--Gamma density with parameters (α, ρ, τ) for $\rho > 1$ and $\alpha > 2$ given by

$$f(x|n,y) = \frac{\Gamma(\alpha+\rho)}{\Gamma(\alpha)\Gamma(\rho)} \frac{x^{\alpha-1}\tau^\rho}{(\tau+x)^{\alpha+\rho}}$$

where $\rho = \rho_0 + n\alpha$ and $\tau = \tau_0 + y$. In the presence of zero initial stock and negligible fixed ordering costs, the optimal stocking quantity S^* solves

$$\frac{C_u}{C_u + C_0} = B\left(\frac{S^*}{S^*+\tau}, \alpha, \rho\right) + \frac{S^*\tau}{(\tau+Q)^2} \times b\left(\frac{S^*}{S^*+\tau}, \alpha, \rho\right)\left(1 - \frac{\alpha+1}{\tau\alpha}\right)$$

where $B(x, v, w)$ and $b(x, v, w)$ are the cdf and pdf of a beta random variable with parameters v and w.

In the presence of censored data, the Bayesian analysis of inventory models is quite limited. The only exact analysis is by Lariviere and Porteus [12] where the demand has the newsvendor distribution of the form

$$f(x|\theta) = \theta d'(x)\exp(\theta d(x))$$

where $d(x)$ is a positive, differentiable, and increasing function with derivative $d'(\cdot)$. If X has a newsvendor distribution, then $d(X)$ has an exponential distribution with rate θ. The gamma distribution is a conjugate prior for all newsvendor distributions. When θ has a gamma prior with initial shape and scale parameters α_0 and β_0, the posterior density of demand, after n observations of sales, is given by

$$f(x|\alpha,\beta) = \frac{\alpha\beta^\alpha d'(x)}{[\beta + d(x)]^{\alpha+1}}$$

where $\beta = \beta_0 + \sum_{i=1}^n d(y_i)$, $\alpha = \alpha_0 + m_n$, y_i denotes the observed sales in period i, and m_n denotes the number of periods without any stockout (i.e., number of uncensored observations). The sufficient statistic for this case is the triplet $(n, m_n, \sum_{i=1}^n d(y_i))$. No other distribution is known to retain its conjugate property with censored data. Berk, Gürler, and Levine [4] propose the use of a two-moment approximation which consists of substituting the exact posterior for the censored observation with another conjugate posterior such that its first two-moments match those of the exact posterior obtained. Suppose the demand is Poisson with unknown parameter λ with (conjugate) gamma prior with parameters α and β. Given any sales observation such that $y = S$, the first two moments of λ are given by

$$m_1 = \frac{1}{(\beta+1)}\left[\alpha + \rho\frac{\frac{\partial}{\partial\rho}A(\rho,S,\alpha)}{A(\rho,S,\alpha)}\right]$$

and

$$m_2 = (\beta+1)^{-2}\left[\alpha(\alpha+1) + \rho(\alpha+3)\frac{\frac{\partial}{\partial\rho}A(\rho,S,\alpha)}{A(\rho,S,\alpha)} + \rho^2\frac{\frac{\partial^2}{\partial\rho^2}A(\rho,S,\alpha)}{A(\rho,S,\alpha)}\right]$$

where $\rho = 1/(\beta + 1)$ and $A(\rho, S, \alpha) = \sum_{i=S}^{\infty} \frac{\Gamma(\alpha+i)}{i!}\rho^i$. Using the gamma posterior of λ with parameters $\alpha^* = \frac{m_1^2}{m_2 - m_1^2}$ and $\beta^* = \frac{m_1}{m_2 - m_1^2}$, the posterior demand distribution is then computed as in the uncensored case.

A similar approximation is developed for the normal demand case with known standard deviation σ and unknown mean μ with a normal prior with initial mean ρ_0 and initial standard deviation τ_0. For the censored observation, the posterior of μ is again normal with parameters ρ and standard deviation τ where

$$\rho = E(\mu \mid y > S) = \rho_0 + \tau_0^S \lambda(S)$$

and

$$\begin{aligned}\tau^2 = (1/\eta^2)\Big(&\rho_0^2\sigma^4 + \eta\tau_0^2\sigma^2 + 2\tau_0^2\sigma^2 L^2\rho_0^2 \\ &+\tau_0^4\eta + \rho_0^2\tau_0^4\Big) + \Big[2\rho_0\tau_0^2\sigma^2 + \tau_0^4(S+\rho_0)\Big] \\ &\times \lambda(S)/\eta - [E(\mu \mid y > S)]^2\end{aligned}$$

where $\lambda(z) = \phi(z)/\overline{\Phi}(z)$ and $\eta = \sigma^2 + \tau_0^2$. The reported numerical results indicate that the approximation is highly satisfactory.

REFERENCES

1. Agrawal, N. and Smith, S. A. (1996). Estimating negative binomial demand for retail inventory management with lost sales. *Nav. Res. Logist.*, **43**, 839–861.
2. Arrow, K. J., Harris, T. and Marschak, J. (1951). Optimal inventory policy. *Econometrica*, **19**, 250–272.
3. Azoury, K. S. (1985). Bayes solution to dynamic inventory models under unknown demand distribution. *Manage. Sci.*, **31**, 1150–1160.
4. Berk, E., Gürler, Ü., and Levine, R. (2001). The Newsboy Problem with Bayesian Updating of the Demand Distribution and Censored Observations. In *Monographs of Official Statistics: Bayesian Methods*, Office for Official Publications of the European Communities, Luxembourg, pp. 21–31.
5. Conrad, S. A. (1976). Sales data and estimation of demand. *Oper. Res. Quart.*, **27**, 123–127.
6. Gallego, M. and Moon, I. (1993). The distribution free newsboy problem: reiew and extensions. *J. Oper. Res. Soc.*, **44**, 825–834.
7. Gupta, A. K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika* , **39**, 260–273.
8. Halperin, M. (1952). Maximum likelihood estimation in truncated samples. *Anns. Math. Stat.*, **23**, 226–238.
9. Hill, R. M. (1997). Applying bayesian methodology with a uniform prior to the single period inventory model. *Eur. J. Oper. Res.*, **98**, 555–562.
10. Iglehart, D. L. (1964). The dynamic inventory problem with unknown demand distribution. *Manage. Sci.*, **10**, 429–440.
11. Johnson, N. L., and Kotz, S. (1969). *Discrete Distributions*. Houghton Mifflin, Boston, Mass.
12. Lariviere, M. A. and Porteus, L. E. (1999). Stalking information: Bayesian inventory management with unobserved lost sales. *Manage. Sci.*, **45**, 346–363.
13. Lowe, T. J., Schwarz, L. B. and McGavin, E. J. (1988). The determination of optimal base-stock inventory policy when the costs of under- and oversupply are uncertain. *Nav. Res. Logist.*, **35**, 539–554.
14. Moon, I. and Choi, S. (1995). The distribution free newsboy problem with balking. *J. Oper. Res. Soc.*, **46**, 537–542.
15. Morton, T. E. (1971). The near-myopic nature of the lagged-proportional-cost inventory problem with lost sales. *Oper. Res.*, **19**, 1708–1716.
16. Nahmias, S. (1994). Demand estimation in lost sales inventory systems. *Nav. Res. Logist.*, **41**, 739–757.
17. Scarf, H. (1958). A Min-max Solution of an Inventory Problem. In *Studies in the Mathematical Theory of Inventory and Production*, K. J. Arrow, S. Karlin, and H. Scarf, eds. Stanford University Press, Stanford, pp. 201–209.
18. Scarf, H. (1959). Bayes solutions of the statistical inventory problem. *Anns. Math. Stat.*, **30**, 490–508.
19. Silver, E. A. (1965). Bayesian determination of the reorder point of a slow moving item. *Oper. Res.*, **13**, 989–997.
20. Zipkin, P. (2000). *Foundations of Inventory Management*. McGraw-Hill, Singapore, pp. 378–385.

FURTHER READING

Lovejoy, W. S. (1990). Myopic policies for some inventory models with uncertain demand distribution. *Manage. Sci.*, **36**, 724–738.

Emre Berk
Ülkü Gürler

NEWTON ITERATION EXTENSIONS

Newton iteration is a powerful method for estimating a set of parameters that maximize a function when the parameters are related to the function nonlinearly. Two general applications of Newton iteration are nonlinear regression (least squares), often referred to as Gauss-Newton iteration, and maximum likelihood estimation*. Since the method is very general, it is usable in many other estimation procedures as well. References containing numerical examples are noted in the bibliography.

Newton iteration is very powerful for many problems and is the most powerful of the gradient procedures, given certain assumptions. (See Crockett and Chernoff [1] and Greenstadt [4].) For other problems, the method has not converged to a maximum or minimum. In addition to describing Newton iteration, this article gives a procedure which can be used to detect troublesome problems and automatically switch to a Newton extension that performs well for a larger class of problems.

NEWTON ITERATION

To apply Newton iteration, the first and second partial derivatives (or at least an approximation to them) are calculated at each iteration. More specifically, for a function $f(\mathbf{X} : \mathbf{b})$ to be maximized, the m-dimensional vector $\mathbf{b}^*$, which maximizes $f(\mathbf{X} : \mathbf{b})$, is calculated by going through a series of iterations with calculated values of $\mathbf{b}$ (i.e., $\mathbf{b}_{(1)}, \mathbf{b}_{(2)}, \ldots$) until $\mathbf{b}^*$ is found. To minimize a function, merely maximize the negative of that function.

If a local maximum separate from the global maximum exists in a region near any of the $\mathbf{b}_{(i)}$, then convergence is likely to be to the local maximum. Saddle points may be readily handled by the extensions to Newton iteration given further on. Since the focus here is on selecting the $\mathbf{b}^*$ that maximizes $f(\mathbf{X} : \mathbf{b})$, in what follows we will simplify notation by dropping the specific recognition of the matrix of variables, if any, in the function to be maximized and write the function as $f(\mathbf{b})$.

Newton iteration is a gradient method of maximization; that is, at each iteration, the next point, $\mathbf{b}_{(i+1)}$, is chosen in the direction of the steepest ascent from the present point, $\mathbf{b}_{(i)}$. The particular concept of distance used in determining steepest ascent is the Newton metric—the $m \times m$ matrix of second partial derivatives of $f(\mathbf{b})$.

The formulas for Newton iteration may be derived by writing out the first three terms of a Taylor expansion about an m-dimensional point $\mathbf{b}_{(i)}$, taking the first partial derivative of $f(\mathbf{b})$, setting the first partial derivative to 0, and solving for $\mathbf{b}$. (See Crockett and Chernoff [1].) The following is obtained:

$$\mathbf{b}_{(i+1)} = \mathbf{b}_{(i)} + \mathscr{L}_{(i)}^{-1}\ell_{(i)},$$

where $\ell_{(i)}$ is the m-dimensional vector of first partial derivatives of $f(\mathbf{b})$ evaluated at $\mathbf{b}_{(i)}$ [i.e., with the m values of $\mathbf{b}_{(i)}$ substituted into the formula for the first partial derivative of $f(\mathbf{b})$] and $-\mathscr{L}_{(i)}$ is the $m \times m$ matrix of second partial derivatives of $f(\mathbf{b})$ evaluated at $\mathbf{b}_{(i)}$.

If the first three terms of the Taylor expansion were sufficiently close to $f(\mathbf{b})$ and if $\mathscr{L}_{(i)}$ were positive definite (i.e., $-\mathscr{L}_{(i)}$ were negative definite) then $\mathbf{b}^*$ would be the maximum of $f(\mathbf{b})$ and this article would be almost complete, saving time for everyone. Since the first three terms do not sufficiently represent $f(\mathbf{b})$, the $\mathbf{b}^*$ which maximizes $f(\mathbf{b})$ must be computed by a series of iterations.

The preceding formula suggests $\mathbf{d}_{(i)} = \mathscr{L}_{(i)}^{-1}\boldsymbol{l}_{(i)}$ as the direction to take at each iteration.

In Newton iteration the length of movement in direction $\mathbf{d}_{(i)}$ is usually generalized so that instead of a step size of one, a step size of $h_{(i)}$ (a scalar) is used, with $h_{(i)}$ varying with each iteration. Thus $\mathbf{b}_{(i+1)}$ with $f(\mathbf{b}_{(i+1)}) > f(\mathbf{b}_{(i)})$ is calculated by the formula

$$\mathbf{b}_{(i+1)} = \mathbf{h}_{(i)}\mathbf{d}_{(i)}.$$