A FORMULA FOR
THE RESERVATION WAGE
AND ITS
LIMITED APPLICABILITY

Coen A.Bernaards

May 24, 1996

Preface

This report was written to complete my study Statistics at the Rijks*universiteit* Groningen and the Masterclass in Applied Statistics of the Mathematical Research Institute. It was developed under the guidance of prof. dr R.D. Gill (University of Utrecht) who also had the privilege of using a data set kindly provided by the CBS. Prof. dr W. Schaafsma read early drafts of the report.

Contents

1 Introduction

1.1 General perspective

In the past decade the Dutch labour market was frequented by a large number of unemployed people. Though the situation improved, the number of job-seekers is still large. Since the mid-eighties the number of job opportunities was increased but unfortunately not everybody took advantage of this development. Many people are still looking for a job, others have averted themselves from the labour market.

The above impression of actual developments at the labour market is very superficial: matters are sketched at one particular moment or compared to other points in time. Individual characteristics, e.g. the frequency of unemployment, or the time to find a new job, are not in the forefront. From a sociological perspective there is a considerable difference between the situation where surplus labour is formed by a small hard core of unemployed and that where a large number of people is involved, each one needing little time to find a job. A 'dynamic' approach to the labour market requires the analysis of the factors involved in the question: who is unemployed and how much time is needed till this person's search is successfull? This kind of analysis indicates the weak groups of society.

The probability for a searcher to find a job is determined by both the probability an employer offers a job and by the probability the searcher accepts it. Individual characteristics determine the former; for the latter the wage is decisive (fringe benefits are not taken into account): the job-seeker accepts the offer if the wage exceeds his so-called "reservation wage". This, of course, is only an idealization because (1) the same person may be confronted with different job offers, (2) it is not possible to express everything in money.

The goal of this research is to examine the relation between the duration of unemployment on the one hand and personal characteristics on the other. Especially the reservation wage and its mathematical implications are subject of investigation. Note that the statistical analysis will have to deal with censored observations. A description of the data will be given in Section 1.2.

Next, in Chapter 2, some theory is presented about the reservation wage. Chapter 3 contains a statistical analysis which does not respect the restrictions imposed by the actual data. In Chapter 4 an attempt is made to say something on the basis of this factual information. Dirty tricks and derivations are relegated to Appendices.

1.2 Description of the data

In this report data from the Socio-Economic Panel (SEP), a study which started in April 1984, is used. The sample for this panel is drawn according to a two-stage sampling scheme. The population consists of k "communities" witk $N_1, \ldots, N_k$ as the number of inhabitants.

From these k communities the ℓ largest ones are included and from the $k - \ell$ other ones $m - \ell$ elements are selected by drawing successively according to probabilities proportional to the size of the community. Next, from the selected communities a sample is drawn such that, in the end, each household has the same success probability. Of each household, every person from sixteen upwards is interviewed (except persons in institutions). From October 1985 onwards, the SEP consist of about 5000 households.

Interviews for the SEP are conducted twice every year, in October and in April. In each wave the same households participate. Members of households leaving home are automatically added as new households. When a member of the household reaches the age of sixteen this individual is added to the SEP. Some new households are included in addition in order to compensate for panel attrition.

The main interest of this report concerns the duration of unemployment (or search duration). All persons are asked to indicate their main activity in each of the past six months. In the April interview this means last year October till March, in the October interview this means last April up to and including September. note that the main activity during the survey month (April or October) is recorded in the next wave, six months later. From all participants, periods of search from October 1984 onwards are recorded. Also, from newly sampled households all periods of search from October 1984 onwards are asked. It should be noted that individuals participating for the first time are asked about their unemployment during the last twelve months.

Possible answers to the question on main activity are "full time work", "part time work", "looking for work for the first time", "looking for work after loosing a job", "going to school or studying" and "working in own household". Only one activity can be chosen. For the data set, the duration of unemployment is computed for people without a job.

For this report five SEP waves are available; the ones recorded in April 1985 up to and including October 1987. Using the question about main activity, duration of unemployment is derived in months. At the start of a job-searching period some background variables are recorded like age and sex. Furthermore, respondents had to indicate the number of hours they desire to work each week, their unemployment benefit and their reservation wage. This last variable is standardized to forty hours per week in order to end up with comparable quantities.

Questions concerning the reservation wage were recorded only during the waves October 1985 and October 1986. Hence the reservation wage applies to search spells covering either (or both) of these months.

Quite a number of spells ended in September 1985 or September 1986 while the person concerned reported to be unemployed in October (recall that the main activity in October is investigated six months later, in March). Probably the search for work was not successfull in September and the main activity in October should have been "looking for work", perhaps this was reported due to a lack of memory to distinguish six months later between October and September. The conclusion is that questions concerning the reservation wage also apply to these spells; they too, have been selected in the analysis though the calculation of duration of unemployment was not corrected for this phenomenom.

Some attention should be paid to the quality of the variables reservation wage and benefit. Eight repondents report a reservation wage of less than 400 guilders, thirteen

report a benefit less than 400 guilders. These observations are disregarded. Four individuals reported their reservation wage lower than their benefit. They are also not taken into account.

Lastly, one should realize that a person can become unemployed more often than once, during the considered time span. Search spells originating from the same individual are contained in the data.

The reduction has the effect that the data consists of only 327 entries. Notice that there is selection mechanism; search durations falling between October 1985 and September 1986 are removed from the analysis while the remaining periods have on average a longer duration.

The following variables were recorded.

- starting month of unemployment;
- closing month of the study or of the period of unemployment;
- a censoring variable δ (0 if unemployment ended before the study, 1 if the individual in question was still unemployed at the closing date October 1987);
- reservation wage (r);
- unemployment benefit (b);
- a vector of covariables $(\mathbf{Z})$ are available. At the moment an individual starts looking for a job the following covariates are recorded:
  - whether or not the individual is married, a 0-1 variable, 1 when married,
  - the place of the individual in the household, a 0-1 variable, 1 when person is either breadwinner or single,
  - the level of education, a 0-1 variable, 1 when educational level is secondary vocational training, 0 when lower educational level,
  - the person's age measured in years,
  - the size of the person's residence, a 0-1 variable, 1 when living in a town of $\geq$ 100.000 inhabitants,
  - the person's sex, a 0-1 variable, 1 when individual is female,
  - whether or not the individual did have a job before, a 0-1 variable, 1 when having worked before, 0 when waiting for the first job.

2 Model building

2.1 The model

At the labour market jobs are looking for workers and workers are looking for jobs. If some person enters the market then he is entitled to a benefit of, say, b money units per unit of time. Jobs will be offered to him at time points $t^1, t^2, \ldots$ ($0 =$ time of entry) which will be distributed according to a Poisson process with rate λ. This assumption implies that inter-arrival times are independent and exponentially distributed with parameter λ, i.e. with density $\lambda e^{-\lambda t}$ ($t \geq 0$). Every job offer is accompanied by a wage. The wages $w^1, w^2, \ldots$ are supposed to originate from some distribution with distribution function F, usually lognormal or Pareto. In this research a Pareto distribution is postulated, with distribution function

$$F(w) = \begin{cases} 0 & \text{if } w < c \\ 1 - \left(\frac{c}{w}\right)^{\sigma} & \text{if } w \geq c. \end{cases} \tag{2.1}$$

and density

$$f(w) = \begin{cases} 0 & \text{if } w < c \\ \sigma c^{\sigma} w^{-(\sigma+1)} & \text{if } w \geq c. \end{cases}$$

In the sequel this distribution will referred to as Pareto(c, σ). Figure 2.1 displays some Pareto densities.

Suppose a job has been offered with wage w. The individual who receives the offer has to choose between two possibilities: acceptance of the job, or rejection (in which case search continues). His decision rule is supposed to be based on a value r, the reservation wage, which depends on λ, b, F and some discount factor ρ. The rule is supposed to be of the form

$$\begin{cases} \text{accept} & \text{if } w \geq r \\ \text{reject} & \text{if } w < r. \end{cases}$$

As a consequence of these assumptions, the reservation wage will satisfy the following equation

$$\frac{\lambda}{\rho} \int_r^{\infty} (w - r) dF(w) = r - b. \tag{2.2}$$

The proof of this relation is relegated to Appendix A. There is exactly one value of r satisfying Equation (2.2) because (1) both sides of (2.2) are continuous functions of r (see Appendix B); (2) the left-hand side is a decreasing function of r and the right-hand side an increasing function of r; (3) if $r = b$ the left-hand side is larger than the right-hand side.
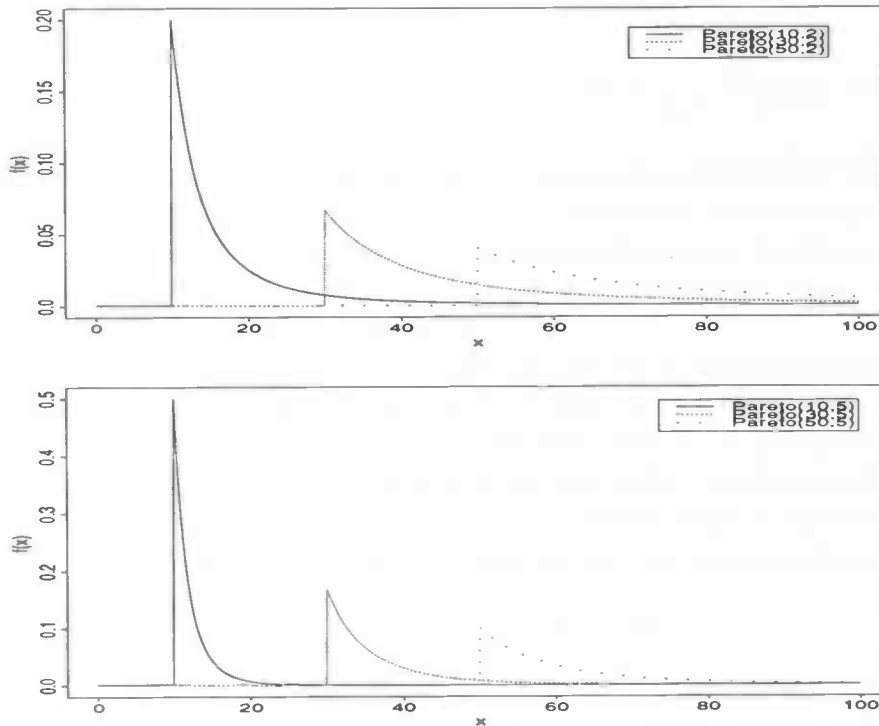
4

Figure 2.1: Pareto densities with scaling factors $\sigma = 2$ and $\sigma = 5$.

Theory on reservation wages can be critized because the assumptions behind (2.2) are questionable, and, more importantly, because the parameters λ, b, c, σ and, possibly, ρ depend on the person who is looking for a job, while this person himself has no knowledge of their precise values. More precisely, λ, b and (c, σ) (or F) depend on 'covariates' or, almost equivalently, on the 'reference class' corresponding to a particular individual. For such a class 'expected duration of unemployment' can be defined in the frequency-theoretic sense and it is in this sense that such population characteristics can be applied to the individuals in such a reference class. Questions concerning the wage distribution F and search time can be discussed in the same way.

It was assumed that the wages follow the Pareto(c, σ) distribution. Substitution of F in Equation (2.2) gives,

$$\frac{\lambda}{\rho} \int_{\max(r,c)}^{\infty} (w - r) \sigma c^{\sigma} w^{-(\sigma+1)} dw = r - b. \tag{2.3}$$

This looks nicer than it is, even if the attention is restricted to a specific reference class with typical values for the covariates. It sounds nice to make assumptions as done earlier and to relate λ, b, c, σ to the reference class. The problem is, however, that the rule (2.2) can only be implemented by the job-seeker if this job-seeker knows λ, c and ρ. As this is an unrealistic supposition, the job-seeker will have to guess. He even will not necessarily

have $r \geq c$. Ignoring this inconvenience by assuming $r \geq c$, (2.3) reduces to

$$\frac{\lambda}{\rho} \int_r^\infty (w - r)\sigma c^\sigma w^{-(\sigma+1)} dw = \frac{\lambda c^\sigma}{\rho(\sigma - 1)r^{\sigma-1}} = r - b. \tag{2.4}$$

Notice that this expression provides an implicity definition of the reservation wage r (unless $b = 0$). Recalling that c is an unknown parameter, what conditions should be imposed to ensure $r \geq c$? All individuals in a reference class face the same c, hence c should be taken not greater than the smallest r of all individuals in the class. Since every individual will set $r \geq b$, c is required to satisfy $b \leq c \leq r$.

Recall the decision if a job is offered: the job is accepted if its wage is equal to or exceeds the person's reservation wage, and rejected in all other cases. Job offers are assumed to arrive according to a Poisson process with rate λ. A person leaves unemployment in the short interval $(t, t+h)$ if and only if the job-seeker receives a new wage offer in the interval, an event of probability λh, and the reservation wage is exceeded, an event of probability $1 - F(r)$. Thus the 'hazard' function for leaving unemployment is $\theta(t) = \theta$ where

$$\theta = \lambda(1 - F(r)), \tag{2.5}$$

Hence the duration of unemployment t is distributed exponentially with parameter θ, $T \sim \mathcal{E}(\theta)$, see, e.g., Kalbfleisch and Prentice (1980). Under the Pareto assumption, Equation (2.5) can be written as

$$\theta = \lambda \left(\frac{c}{r}\right)^\sigma = r^{-\sigma}(\lambda c^\sigma). \tag{2.6}$$

In the sequel it is assumed that σ is a constant. The other parameters c, λ and r, however, will be individual dependent. Hence θ will have the same property.

2.2 The statistical context

An attempt is made to relate the theory of Section 2.1 to the data available. These were reported in Section 1.2. Recall to mind that for n job-seekers outcomes of the following variables were recorded:

- starting month of unemployment;
- closing month of study or of the period of unemployment;
- censoring variable (δ) (0 if unemployment ended before the study, 1 if it extends beyond the closing date October 1987);
- reservation wage (r);
- unemployment benefit (b);
- vector of covariates (**Z**) (see Section 1.2).

By subtracting starting month from the second month number, the (censored) "survival" time $t^* \in \{0, 1, 2, \ldots\}$ is obtained. It describes the duration of unemployment if $\delta = 0$ and the duration until the end of study if $\delta = 1$. The actual survival time (or rather unemployment time) t is equal to t^* if $\delta = 0$ but if $\delta = 1$ the only knowledge is that $t > t^*$. If $\delta = 0$ the time until the study's closing date is also available. This can be useful in the derivation of certain 'reduced-sample' estimates.

For every job-seeker the variables t, r, b and $\mathbf{Z}$ can be discussed though, unfortunately, some observations are censored and many observations are simply missing. The theoretical context should somehow express the effect of r, b and $\mathbf{Z}$ on t. One possibility to do so is by compensating the lack of relevant information about λ and F (or c and σ). This can be done by incorporating some mathematical assumptions. The first one, originating from Jones (1988), is that

$$\log(\lambda c^\sigma) = \mathbf{Z}\beta + u, \tag{2.7}$$

where $\mathbf{Z}$ is the row vector of covariates, β is a vector of regression coefficients and u is an error term which, like $\mathbf{Z}$, will be individual dependent. The distribution from which the error u is taken is supposed to be the same for all individuals. The assumption on the conditional distribution of the survival time, given the individual with its specific vector of covariates $\mathbf{Z}$ and error term u, is the exponential distribution with parameter θ.

What happens when these two assumptions are applied to the theory of Section 2.1? Recall Equation (2.4). It was pointed out several times that this is an exact relation: any parameter, e.g. r, is determined by the other ones. Hence, a testable model arises if it is assumed that the job-seeker implements the reservation wage of Equation (2.4). The model consists of assumption (2.7) substituted into Equation (2.4) and (2.6).

Suppose one is not willing to believe that the reported reservation wage will coincide the choice of r determined by Equation (2.4). In that case it is useful to distinguish between the choice of r determined by Equation (2.4) and the reservation wage actually reported. The r from (2.4) is deterministic. It is assumed that the reservation wage R reported, is a random variable and that

$$R = r + \zeta \tag{2.8}$$

where ζ is an error-term having mean zero and dispersion parameter τ, e.g. a variance. Recall to mind that $T \sim \mathcal{E}(\theta)$ implies $\theta T \sim \mathcal{E}(1)$. Now substitute R for r into (2.6) and use assumption (2.7), then the duration of unemployment t of a particular individual is composed as follows

$$t = r^\sigma e^{-(z\beta + u)} \tilde{e} \tag{2.9}$$

where $\tilde{e}$ is the outcome of $\mathcal{E}(1)$, u is a random disturbance and r is the outcome of R defined in (2.8) (possibly with $\zeta = 0$). Of course the random variables whose distribution is not yet discussed is postulated to follow a normal distribution. As a consequence, when r the outcome of R, is substituted into Equation (2.4) the following relation has to hold

$$\mathbf{Z}\beta + u = \log\left(\rho(\sigma - 1)r^{\sigma - 1}(r - b)\right). \tag{2.10}$$

When the reported reservation wage is modelled according to (2.8) a model arises consisting of (2.8), (2.9) and (2.10).

2.3   Recapitulating the models

In the preceding sections several ideas were developed with respect to the reservation wage, the duration of unemployment, etc. These ideas have to be implemented in the form of a statistical model such that the available data can be evaluated. This leads to the idea that the model has to be sufficiently simple: parameters have to be identifiable, etc. On the other hand it should not ignore basic features of the way the job-seeking process is thought of.

Recall from Section 1.2 that the data actually available are very fragmentary. This may hamper that implementation of any model but it will certainly lead to unresolvable difficulties if the model is not sufficiently simple. In the present section where an overview is given of some models, the fragmentary charcter of the data will be ignored. All models are based on (1) the idea that the duration of unemployment has an exponential distribution, given certain covariable valuesand error terms, and (2) the idea that wage offers follow a Pareto distribution, again given the covariable values and certain error terms. The relation (2.4) or some slight modification of it is always crucial.

Model 1

In this model the main assumption is that there is no distinction between the theoretical reservation wage satisfying Equation (2.4) and the one reported reported by the job-seeker. Also Equation (2.10) has to be valid. The model consists of the following relations

$$
\begin{aligned}
t &= r^{\sigma} e^{-(z\beta + u)} \tilde{e}, \\
z\beta + u &= \log\left(\rho(\sigma - 1)r^{\sigma - 1}(r - b)\right).
\end{aligned}
$$

where r is the reservation wage satisfying (2.4) and $\tilde{e}$ follows an $\mathcal{E}(1)$ distribution. Table (2.1) gives an overview of the variables in the model. In Section 3.1 a method for

Observed variables	t, z, r, b
Latent variables	$u, \tilde{e}$
Random variables	$u, \tilde{e}$
Parameters	β, σ
Postulated constant	ρ

Table 2.1: Type of variables

parameter estimation will be proposed.

Model 2

This model compensates for the fact that the job-seeker will never report a reservation wage equal to the theoretical one of Equation (2.4). A random variable R, defined as

in (2.8), representes the reported reservation wage. The outcome r of R still has to satisfy Equation (2.10). The model consists of the following relations

$$
\begin{aligned}
t &= r^\sigma e^{-(z\beta+u)}\tilde{e}, \\
z\beta + u &= \log\left(\rho(\sigma-1)r^{\sigma-1}(r-b)\right), \\
R &= r + \zeta.
\end{aligned}
$$

Table 2.2 gives an overview of the variables. The variable $\tilde{e}$ follows the $\mathcal{E}(1)$ distribution.

Observed variables	t, z, b, r
Latent variables	u, $\tilde{e}$, ζ
Random variables	R, u, $\tilde{e}$, ζ
Parameters	β, σ
Postulated constant	ρ

Table 2.2: Type of variables

2.4 Gorter's model

Before proceeding with the estimation of parameters, some remarks should be made on the model used by Gorter (1991), the report from which this research originated. The present author has to admit the analysis is not fully understood and that is why the reader could also be advised to skip to Chapter 3. Gorter's model consists of

$$
\mathbf{Z}\beta + u = \log\left(\rho(\sigma-1)r^{\sigma-1}(r-b)\right)
$$

and of the expectation of the logarithm of

$$
\theta = r^{-\sigma}e^{\mathbf{Z}\beta+u}.
$$

Rewriting (2.11) in the form

$$
\mathbf{Z}\beta + u - \log\rho(\sigma-1) - (\sigma-1)\log r = \log(r-b) \tag{2.11}
$$

the idea is to use a Taylor expansion on the right-hand side:

$$
\log(r-b) = \delta_0 + \delta_1 \log r + \delta_2 \log b + v, \tag{2.12}
$$

where v represents the higher order terms. Notice that the coefficients δ can be otained by partial differentiation, e.g. $\delta_2 = r/(r-b)$. Substitute (2.12) into (2.11) and rewrite this to

$$
\begin{aligned}
\log r &= \frac{1}{\delta_2 + \sigma - 1}\left[-\delta_0 - \log\rho(\sigma-1) - \delta_2 \log b + \mathbf{Z}\beta + u - v\right] \\
\Leftrightarrow \log r &= \text{constant} + \mathbf{X}\pi + \mathbf{Z}\alpha + \varepsilon,
\end{aligned}
$$

9

where $\mathbf{X} = (\log b, \log c)'$ and the error term ε equals the sum of the error term u and the remainder of the Taylor expansion v. A problem is that ε is correlated with u.

The second equation of his model derives from Equation (2.11). He suggests that this relation naturally suggests a logarithmic specification. Consider the expectation of the logarithm of elapsed duration,

$$E(\log t | r, \mathbf{Z}) = -\gamma + \sigma \log r - \mathbf{Z}\beta - E(u | r, \mathbf{Z}), \qquad (2.13)$$

where γ is Euler's constant. The integral of this expectation $\int_0^\infty e^{-\mu t} \log t \, dt$ can be found in Ryshik and Gradstein (1963). Assume $E(u|.) = 0$. The simultaneous equations model used by Gorter is,

$$\begin{aligned} \log r &= \text{constant} + \mathbf{X}\pi + \mathbf{Z}\alpha + \varepsilon, \\ \log t &= -\gamma + \sigma \log r - \mathbf{Z}\beta - u, \end{aligned} \qquad (2.14)$$

which applies to every individual of which the variable end month of unemployment was not censored.

While developing the Equations of (2.14), quite a number of difficulties presented themselves. If not enough, the data revealed censored observations. Gorter (1991) used a method called two stage least squares, see e.g. Stewart (1991) or Johnston (1984); a method which does not allow use of censored observations. Furthermore, one has to make use of approximations.

10

3 Statistical analysis: theory

In Section 2.3 two models were developed. The present chapter is concerned with a theoretical analysis. Recall from Chapter 1 that one of the goals of this report is that methods of estimation should not suffer from censored observations.

For Model 1 of Section 2.3 parameters may be estimated via goodness of fit. In Section 3.1 this method is elaborated.

The present section does, however, not elaborate on Model 2 of Section 2.3. The parameters may be estimated using a method called Gibbs sampling. An introduction to Gibbs sampling is the subject of Appendix C. Some remarks on its application to the underlying model are relegated to Appendix D.

3.1 Testing the goodness of fit

Recall Model 1 from Section 2.3:

$$t_i = r_i^{\sigma} e^{-(z_i\beta + u_i)} \tilde{e}_i, \tag{3.1}$$

$$z_i\beta + u_i = \log\left(\rho(\sigma - 1)r_i^{\sigma-1}(r_i - b_i)\right). \tag{3.2}$$

Here the subscript i ($i = 1, \ldots, n$) refers to the i-th individual which, incidentally, has z_i as its (row) vector of covariates. The values ρ, r_i, b_i and Z_i are supposed to be known. The unknown parameter σ is supposed to be same for all individuals.

When the reservation wage was discussed, in Section 2.1 and in Appendix A, attention was drawn that quite a number of assumptions were at the basis of it. In order to get an impression of the tenability of the theory, the goodness of fit of the exponential distribution can be tested. A goodness of fit procedure is a statistical test of a hypothesis that the sampled population has a certain property. Define

$$y_i = \log\left(\rho(\sigma - 1)r_i^{\sigma-1}(r_i - b_i)\right).$$

If σ is postulated, y_i is observable. Using assumption (2.7) and Equation (3.1), β can be estimated using the method of least squares.

In order to test the model, including σ, θ_i can be used because $T_i \sim \mathcal{E}(\theta_i)$. If there were no censoring (i.e. $v_i = \infty$), replace $\theta_i t_i^* = \theta_i t_i$ by $e^{-\theta_i t_i} = p_i$ which, under the hypothesis, is a sample from a uniform distribution on $[0, 1]$. Among others, the test by Neyman (1937) applies.

In the case of censored observations, the distribution of T_i is truncated at $\theta_i v_i$ as in Figure 3.1. Take all observations with $\theta_i v_i > k$, for k fixed, and test whether the at k truncated observations fit the exponential distribution. Note that the test consists of two
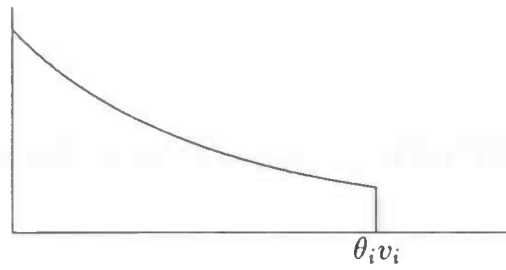
Figure 3.1: Density of an exponential distribution truncated at $\theta_i v_i$

parts. First, the empirical distribution function $F_n(c)$ must fit $P(T_i < c) = 1 - e^{-c}$. Second, the distribution of the spells given $T_i < k$ is

$$F(t) = \frac{1 - e^{-t}}{F(k)}. \tag{3.3}$$

The m observations satisfying $t_i < k$ are distributed according to (3.3). This gives a sample of size m from a uniform distribution on the unit interval. The same tests apply as for the uncensored observations.

4 Statistical analysis: practice

4.1 Results

The present author had hoped to apply the theory of Chapter 3 but, unfortunately, the data available is not satisfactory. Quite a number of the observations of the form t_i, v_i, r_i, b_i, z_i $(i = 1, \ldots, n)$ are missing, especially from r_i and b_i. If observations are missing, then the reason may often be of a non-random character and biased results will appear if proceeded as if the scores were missing at random. This problem can be avoided if one first conditions on the covariables Z_i. However, the number of missing observations is quite large and inference would be based on a small number of individuals. Table 4.1 gives an impression of a few variables.

Variable	# obs.	mean	st.dev.
r	81	1653	615
b	83	1055	329
$r - b$	43	678	492
t with censored obs.	327	6.9	4.8
t without censored obs.	248	5.9	5.4

Table 4.1: Overview of some variables

4.2 Conclusion

It was pointed out several times that the weaknesses of the equation for the reservation wage are its underlying assumptions. The method of goodness of fit, applicable to Model 1 of Section 2.3 and proposed in Section 3.1, can give a hand in order to determine how well the theory fits the data and for which values of σ the theory may be valid. This may serve as a start before any further analysis is carried out. Yet a disadvantage of this method is that σ has to be postulated instead of estimated. Methods of direct estimation are available, however. Gorter (1991) uses two-stage least squares which has the disadvantage that it does not provide the means to use censored observations. Other methods can be sought in the atmosphere of survival analysis; Kortram et al. (1995) may serve as an introduction. Problems concerning missing observations are discussed in Leunis and Altena (1996). The present author, however, recommends to test goodness of fit before proceeding with any further analysis.

For Model 2 of Section 2.3 Gibbs sampling may serve as a start for paramater estimation. Presently the author cannot overlook the quality of the use of this method. Further research is needed in this direction.

A   The reservation wage

An optimal policy for either accepting or rejecting a job offer will be given though the policy cannot be regarded as more than an idealization, a theoretical contruction. The underlying assumption that the job-seeker knows his λ, F and ρ is the weak point of this theory. Let ρ be a discount factor such that one unit of money at time t has a value of $e^{-\rho t}$ now. Assume that offers arrive according to a Poisson process with rate λ, inter-arrival times are distributed exponentially with parameter λ. The benefit per time unit (now) is denoted as b and supposed to be constant in time and t denotes the outcome of T, the seeking time from now ($t = 0$) till the first time a job offer appears. Let I be the expected future income (at $t = 0$).

Suppose at time t an offer is made implying a wage w, obtained by drawing from some distribution F. The enjoyed benefit till that time is worth now

$$\int_0^t b e^{-\rho s} ds.$$

Moreover, the return of the offer equals

$$\begin{cases} \int_t^\infty w e^{-\rho s} ds = \frac{w}{\rho} e^{-\rho t} & \text{if the offer is accepted,} \\ I e^{-\rho t} & \text{if the offer is rejected.} \end{cases}$$

This implies that a job is accepted by a rational person who 'knows everything', if $\frac{w}{\rho} > I$ or, equivalently, if $w > r$ where $r = \rho I$ is the reservation wage. The size of the wage offer W and the seeking time T are assumed independent though this questionable. Then the expected revenue now is

$$\begin{aligned} I &= E\left(\int_0^T b e^{-\rho s} ds + \max\left\{ \frac{W}{\rho}, I \right\} e^{-\rho T} \right) \\ &= \frac{b}{\rho} + E\left[\left(\max\left\{ \frac{W}{\rho}, I \right\} - \frac{b}{\rho} \right) e^{-\rho T} \right] \\ &= \frac{b}{\rho} + \left(E \max\left\{ \frac{W}{\rho}, I \right\} - \frac{b}{\rho} \right) E e^{-\rho T} \\ &= \frac{b}{\rho} + \left(E \max\left\{ \frac{W}{\rho}, I \right\} - \frac{b}{\rho} \right) \frac{\lambda}{\lambda + \rho}. \end{aligned} \tag{A.1}$$

This expression is computed as follows. Note that

$$\begin{aligned} P\left(\max\{ \rho^{-1} W, I \} < c \right) &= \begin{cases} 0 & \text{if } c \leq I \\ P(W < \rho c) & \text{if } c > I \end{cases} \\ &= \begin{cases} 0 & \text{if } c \leq I \\ F(\rho c) & \text{if } c > I \end{cases} \end{aligned}$$

15

Next the expectation becomes

$$
\begin{aligned}
E \max\left\{\rho^{-1}W, I\right\} &= IF(\rho I) + \int_I^\infty cdF(\rho c) \\
&= IF(r) + \rho^{-1}\int_r^\infty wdF(w) \\
&= I - I(1 - F(r)) + \rho^{-1}\int_r^\infty wdF(w) \\
&= I + \rho^{-1}\int_r^\infty (w - r)dF(w)
\end{aligned}
\tag{A.2}
$$

Substitution of (A.2) into (A.1) gives

$$
I = \frac{b}{\rho} + \left(I + \frac{1}{\rho}\int_r^\infty (w - r)dF(w) - \frac{b}{\rho}\right)\frac{\lambda}{\lambda + \rho}
$$

or, equivalently,

$$
I\left(\frac{\rho}{\lambda + \rho}\right) - \frac{b}{\rho}\left(\frac{\rho}{\lambda + \rho}\right) = \frac{1}{\lambda + \rho}\frac{\lambda}{\rho}\int_r^\infty (w - r)dF(w)
$$

which reduces to

$$
I\rho - b = \frac{\lambda}{\rho}\int_r^\infty (w - r)dF(w)
$$

and, hence

$$
r - b = \frac{\lambda}{\rho}\int_r^\infty (w - r)dF(w),
$$

which is exactly Equation (2.2).

B   Some continuity properties

To establish the continuity of $\int_r^\infty (w - r)dF(w)$ as a function of r it is required that the expected wage $\int_r^\infty w\,dF(w)$ is finite. Next the argument is that

$$\int_r^\infty (w - r)dF(w) = \int_0^\infty (w - r)\mathrm{I}_{(r,\infty)}(w)dF(w)$$

where $(w - r)\mathrm{I}_{(r,\infty)}(w)$ is a continuous function of r. In fact it is a uniformly continuous function and, hence, the integral is continuous (because $\int_0^\infty dF(w) < \infty$).

Another argument is that if $r_n \to r$ the functions $(w - r_n)\mathrm{I}_{(r_n,\infty)}(w)$ converge pointwise to $(w - r)\mathrm{I}_{(r,\infty)}(w)$ while $w\mathrm{I}_{[0,\infty)}(w)$ is an integrable upper bound.

C   Introduction to Gibbs sampling

Several articles and books give an introduction to a method called Gibbs sampling. Elementary introductions are given in Casella and George (1992) and Tanner (1993). The introdution used here originates from Gelfand et al. (1990). Remarks on censoring are from Smith and Roberts (1993), Gelfand et al. (1992).

C.1   Stating the problem

In Bayesian statistics parameters are regarded as outcomes of random variables the distributions of which are postulated. As a consequence, parameters observed and latent random variables are all regarded as random variables and the joint distribution is regarded as known though it is more or less a matter of postulating the density $f(u_1, \ldots, u_n)$. If one is interested in a particular parameter, say u_1, then one will first have to incorporate the information available. This can be modelled by regarding $u_{k+1}, \ldots, u_n$ as observed values. The effect is then that the conditional density

$$f(u_1, \ldots, u_k) = \frac{f(u_1, \ldots, u_n)}{\int \cdots \int f(u_1, \ldots, u_n) du_1 \cdots du_k}$$

is considered for fixed observed values $u_{k+1}, \ldots, u_n$. Suppose that such posterior density is available and the marginal density of the first coordinate is of interest

$$f(u_1) = \int \cdots \int f(u_1, u_2, \ldots, u_k) du_2 \cdots du_k. \tag{C.1}$$

Calculation of (C.1) might not always be possible since a primitive might not exist or the dimension is simply too large. The Gibbs sampler allows to generate a sample $U_{11}, \ldots, U_{1m} \sim f_1(u_1)$ without requiring f_1. By taking the sample large enough, characteristics of $f_1(\cdot)$ can be calculated to the desired degree of accuracy.

C.2   The Gibbs sampler

For a collection of random variables with density $f(u_1, u_2, \ldots, u_k)$, the full conditional density of u_s, given the remaining variables, is denoted by $f_s(u_s | u_r, r \neq s)$, $s = 1, \ldots, k$ and the marginal density is denoted by f_s, $s = 1, 2, \ldots, k$.

Consider the following problem. Suppose it is possible to draw random samples of U_s from $f_s(u_s | u_r, r \neq s)$ for specified values of u_r, $r \neq s$, $s = 1, 2, \ldots, k$. Can an iterative scheme be found that enables one to make sample-based estimates, $\hat{f}_s$, say of the marginal densities $f_s(\cdot)$, $s = 1, \ldots, k$? The question can be solved as follows. Given an initial set of values $u_1^{(0)}, \ldots, u_k^{(0)}$, draw $U_1^{(1)}$ from $f_1(u_1 | u_2^{(0)}, \ldots, u_k^{(0)})$, then $U_2^{(1)}$ from

$f_2(u_2|u_1^{(1)}, u_3^{(0)}, \ldots, u_k^{(0)})$, and so on up to $U_k^{(1)}$ from $f_k(u_k|u_1^{(1)}, \ldots, u_{k-1}^{(1)})$ to complete one iteration of the scheme. After t such iterations one arrives at a joint sample $(U_1^{(t)}, \ldots, U_k^{(t)})$. Geman and Geman (1984) showed that under mild conditions $(U_1^{(t)}, \ldots, U_k^{(t)}) \overset{d}{\to} (U_1, \ldots, U_k) \sim f(u_1, \ldots, u_k)$ as $t \to \infty$. Hence for t large enoug, $U_s^{(t)}$ for example can be regarded as a sample variate from $f_s(u_s)$. If this process if replicated in parallel m times, iid k-tuples $(U_{1j}^{(t)}, \ldots, U_{kj}^{(t)}), j = 1, 2, \ldots, m$ result.

Inference for U_s, like center and dispersion, can be calculated directly via the $U_{sj}^{(t)}$. A kernel density estimate for $f(u_s)$ based on the $U_{sj}^{(t)}$ can be obtained readily and should be adequate if the number of replications, m, is large enough.

In the Bayesian context the U_i are the unknown model parameters (or possibly unobserved data) of interest. All distributions are viewed as conditional on the observed data, thus the marginal densities, $f(u_s)$, become the desired marginal posterior distributions of the parameters (or unobserved data).

So far as ease of drawing samples from the complete conditional distributions is concerned, in many cases the likelihood and prior forms specified in the Bayesian model lead to familiar standard full conditional forms, such as Normals and Gammas, and implementation is immediate.

C.3   Censoring in the Gibbs sampler

When the problem at hand contains censored observations, the missing data should be reintroduced as further unknowns, additional to the unknown model parameters, into the Gibbs sampler. Let $z = (y_1, \ldots, y_s, y_{s+1}, \ldots, y_n)$ denote the intended data but only $y = (y_1, \ldots, y_s)$ has been observed, the $y' = (y_{s+1}, \ldots, y_n)$ are subject to right censoring with outcomes $V_j \geq y_j, j = s+1, \ldots, n$. Under the underlying censoring mechanism the posterior will be of the form,

$$f(u_1, .., u_k|V, y) \propto \prod_{i=1}^{s} f(y_i|u_1, .., u_k) \prod_{j=s+1}^{n} \int_{V_j}^{\infty} f(y_j|u_1, .., u_k)dy_j f(u_1, .., u_k). \qquad \text{(C.2)}$$

with $V_j = (V_{s+1}, \ldots, V_n)$. Typically, the unkown model parameters $U_1, \ldots, U_k$, given by $f(u_1, \ldots, u_k|y)$ will be messy, whereas the intended posterior $f(u_1, \ldots, u_k|z)$ will be a tractable form. So, instead of deriving the full conditionals from (C.2), the censored observations are introduced as further unknowns, the corresponding full conditionals have the form,

$$f(u_1, \ldots, u_k|V, y, y') = f(u_1, \ldots, u_k|y, y'),$$
$$f(y'|V, u_1, \ldots, u_k, y) = f(y'|u_1, \ldots, u_k, V) = \prod_{j=s+1}^{n} \int_{V_j}^{\infty} f(y_j|u_1, \ldots, u_k)dy_j.$$

The first of these is just the joint posterior which would have appeared if there had been no censoring. The second is just the joint distribution of the censored observations, given $u_1, \ldots, u_k$. Random variate generation from the former will typically be straightforward; generation from the latter reduces to sampling from the truncated distribution.

19

D Estimation via Gibbs sampling

The original goal of the present section was to develop a Gibbs sampler for estimating of the parameters β and σ, using Equations (3.1) and (3.2). The idea for estimating β used here originates from Gelfand et al. (1990). Other examples of Gibbs sampling are Baks (1995) and Gelfand et al. (1992). A review of Gibbs sampling is given in Appendix C.

The Gibbs sampler without censoring

Apart from the unknown parameter σ, the right-hand side

$$y_i = \log\left(\rho(\sigma - 1)r_i^{\sigma-1}(r_i - b_i)\right) \quad i = 1, \ldots, n \tag{D.1}$$

of (3.1) is observable. Then (3.1) can be interpreted as a model of linear regression

$$\mathbf{y} = \mathbf{Z}\beta + u \tag{D.2}$$

where $\mathbf{y}$ is a n-dimensional column vector and $\mathbf{Z}$ is a $n \times 7$ matrix defined as

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \\ \vdots \\ \mathbf{Z}_n \end{bmatrix}.$$

For the sake of convenience it will be assumed that $u_1, \ldots, u_n$ in (D.2) are outcomes of independent random variables $U_1, \ldots, U_n$ having the $\mathcal{N}(0, 1/\tau)$ distribution.

If σ would be known then the usual theory of least squares applies, see e.g. Hendriks (1996). Equation (3.2) is exploited to obtain information about σ.

Recall that the uniformly minimum variance unbiased estimator for β in the linear model $\mathbf{y} = \mathbf{Z}\beta + u$, equals $\hat{\beta} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$ with u_i iid $\mathcal{N}(0, 1/\tau)$. The dispersion matrix of $\hat{\beta}$ is given by $(1/\tau)(\mathbf{Z}'\mathbf{Z})^{-1}$.

Assume β originates from a seven-dimensional (the number of covariates observed on one individual) multivariate normal distribution $\mathcal{N}_7(\mu, \Sigma)$. This is exactly the weak point of this method. When it is possible to estimate the vector β via an UMVU estimator, it is undesirable and unnecessary to assume some multivariate distribution. The present author does not believe that Gibbs sampling is the answer to the question of estimation in the underlying problem.

Under the multivariate normal assumption, the question of interest on mean and dispersion of β moves towards estimation of μ and Σ. A hierchical Bayesian approach requires specification of the priors for τ, μ and Σ. The priors are assumed independent and to have

a Normal-Wishart-Gamma form,

$$
\begin{aligned}
p(\mu) &\sim \mathcal{N}(\eta, C), \\
p(\Sigma^{-1}) &\sim \mathcal{W}((sR)^{-1}, s), \\
p(\tau) &\sim \mathcal{G}(v_0, v_0 w_0).
\end{aligned}
\tag{D.3}
$$

For determination of the posterior distribution of μ consider its estimator $\hat{\beta}$. Then, by Theorem F.3, the posterior distribution of μ equals,

$$
p(\mu | \hat{\beta}, \tau, \Sigma^{-1}, \sigma, \text{DATA}) \sim \mathcal{N}((\Sigma^{-1} + C^{-1})^{-1}(\Sigma^{-1}\hat{\beta} + C^{-1}\eta), (\Sigma^{-1} + C^{-1})^{-1}). \tag{D.4}
$$

Now assume that the μ sampled from (D.4) is the exact μ from $\beta \sim \mathcal{N}(\mu, \Sigma)$. Under this assumption theory on multivariate analysis (e.g. Theorem 7.2.2 of Anderson (1984)) provides $[(\beta - \mu)(\beta - \mu)']^{-1} \sim \mathcal{W}(\Sigma^{-1}, 1)$. This result and the prior (D.3) for Σ^{-1} give the posterior of Σ^{-1} when applied to Theorem F.4,

$$
p(\Sigma^{-1} | \mu, \hat{\beta}, \tau, \sigma, \text{DATA}) \sim \mathcal{W}\left(\left[sR + (\hat{\beta} - \mu)(\hat{\beta} - \mu)' \right]^{-1}, s + 1 \right). \tag{D.5}
$$

Remains the precision τ. The model $\mathbf{y} \sim \mathcal{N}(\mathbf{Z}\beta, 1/\tau)$ consists of exactly n (the dimension of $\mathbf{Z}\beta$) normals with precision τ. The prior (D.3) applied to Theorem F.5 results in the posterior

$$
p(\tau | \Sigma^{-1}, \mu, \hat{\beta}, \sigma, \mathbf{y}, \text{DATA}) \sim \mathcal{G}\left(v_0 + \frac{n}{2}, v_0 w_0 + \frac{1}{2}(\mathbf{y} - \mathbf{Z}\hat{\beta})'(\mathbf{y} - \mathbf{Z}\hat{\beta}) \right). \tag{D.6}
$$

For the estimation of β the posteriors have now been deduced expect for the hyperparameters η, C, s, R, v_0 and w_0; they are left for specification of information provided by the data.

Recall from Equation (D.1), y_i was observed apart from σ, the scaling parameter of the Pareto(c, σ) distribution. How to estimate σ using the Gibbs sampler?

As a start, try reasoning straightforward. To make life more pleasant, take $f_i(w_i | c, \sigma) \sim$ Pareto(c, σ) for $i = 1, \ldots, n$ (i.e. all individuals face the same parameter c). Then the joint distribution equals,

$$
\begin{aligned}
p(w_1, \ldots, w_n | c, \sigma) &= \prod_{i=1}^{n} \sigma c^{\sigma} w_i^{-(\sigma+1)} I_{[c,\infty)}(w_i) \\
&= \sigma^n \left(\prod_{i=1}^{n} c \right)^{\sigma} \left(\prod_{i=1}^{n} w_i \right)^{-(\sigma+1)} I_{[c,\infty)}(w_i).
\end{aligned}
$$

Following Zellner (1971), a noninformative prior is assumed for σ, $f(\sigma) \propto 1/\sigma, 0 < \sigma < \infty$. This results in the posterior

$$
p(\sigma | c, w_1, \ldots, w_n) \sim \mathcal{G}\left(n, \frac{\log(w_1, \ldots, w_n)}{nc^n} \right).
$$

21

Unfortunately, this distribution depends on the unobserved quantities $w_1, \ldots, w_n$, the wages accepted by the different job-seekers, and c. Despite its nice form, this posterior isn't of much help.

Here's another approach to the estimation of σ. According to Equation (3.2), the durations of unemployment, t_i, are distributed exponentially with parameter

$$\theta_i = r_i^{-\sigma} e^{y_i}, \quad i = 1, \ldots, n. \tag{D.7}$$

Suppose θ_i is given the prior $\mathcal{G}(p_{i0}, q_{i0})$. Then, by Theorem F.6, the posterior $p(t_i|\theta_i)$ is a $\mathcal{G}(p_{i0} + 1, q_{i0} + t_i)$ distribution. When a draw from this distribution replaces θ_i in (D.7), σ can be solved in this equation for the ith individual.

The overall value of value of σ is estimated by averaging all σ's solved as mentioned above.

Censoring in the Gibbs sampler

Recall that one of the targets of this report is to develop methods of estimation insensitive to censored observations. However, when the Gibbs sampler was specified for β and σ, observations were assumed to be observed completely. Presently, the aim is to extend the Gibbs sampler to the case of right censored observations. The concept of the use of censored observations in the Gibbs sampler is given in Appendix C.3.

Consider again the n individuals; of some the duration of unemployment has been observed, the remaining ones are right censored. The idea is to introduce the incomplete durations as unknowns in addition to the parameters presented when only uncensored observations were considered. The Gibbs sampler will have an additional part to sample durations for censored observations. The estimation of β and σ will be carried out using both observed durations and sampled ones.

What is the form of this additional part? Recall the exponential parameter for the duration of unemployment (D.7). For β and σ postulated, θ_i only consists of observed quantities. For the censored survival times t_i^* ($\delta_i = 1$), from Section 2.2, the duration t_i is sampled from a trunctated exponential distribution with parameter θ_i, restricted to the range $t_i \geq t_i^*$.

E   On the simulation of some multivariate distributions

For generating Wishart distributed random matrices, the following algorithm, which originates from Odell and Feiveson (1966) was implemented.

Algorithm E.1 *Given a $p \times p$ covariance matrix R in a factored from $R = CC'$, a sample size N, a sequence of independent standardized normal random variates $\{N_{ij}; i = 1, 2, \ldots, p, j = 1, 2, \ldots, p, i < j\}$ and a sequence of independent χ^2 variates $\{V_j, j = 1, 2, \ldots, p\}$ where for each j V_j has $N - j$ degress of freedom; then a sample covariance matrix*

$$S^* = A^*/N \quad where \quad A^* = CB^*C'$$

can be generated by computing the elements b_{ij}^* of the $p \times p$ symmetric matrix B^* where

$$
\begin{aligned}
b_{11}^* &= V_1, \\
b_{jj}^* &= V_j + \sum_{i=1}^{j-1} N_{ij}^2 \quad j = 2, 3, \ldots, p, \\
b_{1j}^* &= N_{1j}\sqrt{V_1}, \\
b_{ij}^* &= N_{ij}\sqrt{V_i} + \sum_{k=1}^{i-1} N_{ki}N_{kj} \quad i < j = 2, 3, \ldots, p.
\end{aligned}
$$

Note that a disadvantage of this algorithm is that every element of the sample covariance matrix has to be computed separately. By symmetry, for a $p \times p$ matrix $p^2 - \frac{1}{2}p(p+1) = \frac{1}{2}p(p-1)$ calculations are necessary. Hence the implementation is of order $O(p^2)$, though independent of the sample size N on which the maximum likelihood estimator R of the covariance matrix is based. The following implementation is for use in S+ and generates a sample of size 1 of a $\mathcal{W}(R, N)$ distribution.

```
rwishart <- function (vmat,ssize,tol = 1e-07){
  p <- ncol(vmat)
  if(max(abs(vmat-t(vmat)))>tol) stop("vmat not symmetric")
  svm<-svd(vmat)
  msqrt<-svm$v %*% (t(svm$u) * sqrt(svm$d))
  N<-matrix(rnorm(p*p),nrow=p)
  V<-rchisq(1,ssize-1)
  W<-diag(p)
  W[1,1]<-V
  for (i in 2:p){
```

```
    V<-cbind(V,rchisq(1,ssize-i))
    W[i,i]<-V[i]+sum((N[1:i-1,i])^2)
    W[1,i]<-sqrt(V[1])*N[1,i]
    W[i,1]<-W[1,i]
    j <- i+1
    while(j<(p+1)){
        W[i,j]<-sqrt(V[i])*N[i,j]+sum(N[1:i-1,i]*N[1:i-1,j])
        W[j,i]<-W[i,j]
        j <- j+1
    }
  }
  ans<- (msqrt %*% W %*% msqrt)/ssize
  return(ans)
}
```

The parameters for this function are a covariance matrix (vmat) and the sample size (ssize). The output is a $p \times p$ matrix. Note that the order $O(p^2)$ is exposed by a while-loop nested within a for-loop. For decomposing $R = CC'$, a *singular value decomposition* is used.

For generating multivariate normals with given mean and covariance matrix, S+ supplies a routine in one of its examples. The idea behind the function is that if $X \sim \mathcal{N}_p(\mu, \Sigma)$ then $(X - \mu)\Sigma^{-1/2} \sim \mathcal{N}_p(0, I_p)$. The following implementation draws a sample of size n from a $\mathcal{N}_p(\mu, \Sigma)$ distribution.

```
rmultnorm <- function(n, mu, vmat, tol = 1e-07){
  p <- ncol(vmat)
  if(length(mu) != p) stop("mu vector is the wrong length")
  if(max(abs(vmat-t(vmat)))>tol) stop("vmat not symmetric")
  vs <- svd(vmat)
  vsqrt <- t(vs$v %*% (t(vs$u) * sqrt(vs$d)))
  ans <- matrix(rnorm(n * p), nrow = n) %*% vsqrt
  ans <- sweep(ans, 2, mu, "+")
  dimnames(ans) <- list(NULL, dimnames(vmat)[[2]])
  return(ans)
}
```

The parameters for this algorithm are the size of the sample to generate (n), the mean vector (mu) and the covariance matrix (vmat). The output is a $n \times p$ matrix where rows correspond to the different samples and the columns to the elements of μ.

F   Theorems needed for the Gibbs sampler

F.1   Some results from matrix algebra

Some proofs will be presented here, needed to establish the Gibbs sampler.

Theorem F.1 *for square, invertible matrices* Φ *and* Σ,

$$\Phi - \Phi(\Phi + N^{-1}\Sigma)^{-1}\Phi = (\Phi^{-1} + N^{-1}\Sigma^{-1})^{-1}.$$

Proof:

$$
\begin{aligned}
\Phi - \Phi(\Phi + N^{-1}\Sigma)^{-1}\Phi &= \Phi(\Phi + N^{-1}\Sigma)^{-1}(\Phi + N^{-1}\Sigma) - \Phi(\Phi + N^{-1}\Sigma)^{-1}\Phi \\
&= \Phi(\Phi + N^{-1}\Sigma)^{-1}(\Phi + N^{-1}\Sigma - \Phi) \\
&= \Phi(\Phi + N^{-1}\Sigma)^{-1}N^{-1}\Sigma \\
&= \left(\left(\Phi(\Phi + N^{-1}\Sigma)^{-1}N^{-1}\Sigma\right)^{-1}\right)^{-1} \\
&= \left(N\Sigma^{-1}(\Phi + N^{-1}\Sigma)\Phi^{-1}\right)^{-1} \\
&= (\Phi^{-1} + N\Sigma^{-1})^{-1}
\end{aligned}
$$

which is the desired result. □

Theorem F.2 *for square, invertible matrices* Φ *and* Σ, *and vectors* x *and* ν,

$$\Phi\left[\Phi + \frac{1}{N}\Sigma\right]^{-1}x + \frac{1}{N}\Sigma\left[\Phi + \frac{1}{N}\Sigma\right]^{-1}\nu = (\Phi^{-1} + N\Sigma^{-1})^{-1}(N\Sigma^{-1}x + \Phi^{-1}\nu).$$

Proof: start by noticing that Theorem **F.1** yields the following two equalities,

$$
\begin{aligned}
\left(\Phi + N^{-1}\Sigma\right)^{-1} &= \Phi^{-1} - \Phi^{-1}\left(\Phi^{-1} + N\Sigma^{-1}\right)^{-1}\Phi^{-1} \\
\left(\Phi + N^{-1}\Sigma\right)^{-1} &= N\Sigma^{-1} - N\Sigma^{-1}\left(\Phi^{-1} + N\Sigma^{-1}\right)^{-1}N\Sigma^{-1}
\end{aligned}
$$

Substibtution into Equation (F.1) and solving gives

$$
\begin{aligned}
&\left(I - \left(\Phi^{-1} + N\Sigma^{-1}\right)^{-1}\Phi^{-1}\right)x + \left(I - \left(\Phi^{-1} + N\Sigma^{-1}\right)^{-1}N\Sigma^{-1}\right)\nu \\
&= x + \nu - \left(\Phi^{-1} + N\Sigma^{-1}\right)^{-1}(\Phi^{-1}x + N\Sigma^{-1}\nu) \\
&= \left(\Phi^{-1} + N\Sigma^{-1}\right)^{-1}\left(\left(\Phi^{-1} + N\Sigma^{-1}\right)(x + \nu) - \Phi^{-1}x - N\Sigma^{-1}\nu\right) \\
&= \left(\Phi^{-1} + N\Sigma^{-1}\right)^{-1}(N\Sigma^{-1}x + \Phi^{-1}\nu)
\end{aligned}
$$

as was to prove. □

F.2 Two theorems on multivariate posterior distributions

To establish the Gibbs sampler in Section D, some posterior distributions of multivariate distributions were needed. Here they are stated.

Theorem F.3 *If $x_1, \ldots, x_N$ are independently distributed, each x_α according to $\mathcal{N}(\mu, \Sigma)$ and if μ has an a priori distribution $\mathcal{N}(\nu, \Phi)$, then the a postreriori distribution of μ given $x_1, \ldots, x_N$ is*

$$
\mathcal{N}\left(\Phi\left(\Phi + \frac{1}{N}\Sigma\right)^{-1}\bar{x} + \frac{1}{N}\Sigma\left(\Phi + \frac{1}{N}\Sigma\right)^{-1}\nu, \Phi - \Phi(\Phi + N^{-1}\Sigma)^{-1}\Phi\right)
$$
$$
\equiv \mathcal{N}\left((\Phi^{-1} + N\Sigma^{-1})^{-1}(N\Sigma^{-1}\bar{x} + \Phi^{-1}\nu), (\Phi^{-1} + N\Sigma^{-1})^{-1}\right).
$$

The proof can be found in Anderson (1984), Theorem 3.4.5. The equivalence follows from Theorems F.1 and F.2.

Theorem F.4 *If A^{-1} has the distribution $\mathcal{W}(\Sigma^{-1}, n)$ and Σ^{-1} has the a priori distribution $\mathcal{W}(\Psi^{-1}, m)$, then the conditional distribution of Σ^{-1} is $\mathcal{W}((A + \Psi)^{-1}, n + m)$.*

This actually is Theorem 7.7.2 in Anderson (1984) though stated in a different way using Theorem 7.7.1 from the same reference.

F.3 Posterior distribution of a Normal sample with known mean and unknown precision

Theorem F.5 *Suppose an iid random sample is available $f(x_1, \ldots, x_n | \tau) \sim \mathcal{N}(\mu, 1/\tau)$, with μ known. Suppose that the precision τ is distributed according to a Gamma distribution $\mathcal{G}(\alpha, \beta)$ with density,*

$$
f(\tau) = \frac{\beta^\alpha}{\Gamma(\alpha)}\tau^{\alpha-1}e^{-\beta\tau}I_{(0,\infty)}(\tau) \tag{F.1}
$$

then the posterior density $f(\tau | x_1, \ldots, x_n)$ is distributed $\mathcal{G}(\alpha + \frac{1}{2}n, \beta + \frac{1}{2}ns^2)$ where $s^2 = (1/n)\sum_{i=1}^n (x_i - \mu)^2$.

Proof: the joint density of $x_1, \ldots, x_n$ looks like,

$$
\begin{aligned}
f(x_1, \ldots, x_n | \sigma^2) &= \prod_{i=1}^n \frac{\sqrt{\tau}}{\sqrt{2\pi}}e^{-\frac{1}{2}\tau(x_i - \mu)^2} \\
&= \left(\frac{\tau}{2\pi}\right)^{\frac{1}{2}n}e^{-\frac{1}{2}\tau\sum_{i=1}^n (x_i - \mu)^2} \\
&= \left(\frac{\tau}{2\pi}\right)^{\frac{1}{2}n}e^{-\frac{1}{2}\tau ns^2},
\end{aligned}
$$

26

since s^2 is sufficient for τ. Then the posterior distribution $f(\tau|x_1, \ldots, x_n)$ can be found by writing down the joint density of $x_1, \ldots, x_n$ and τ and dividing by the marginal distribution of x,

$$
\begin{aligned}
f(\tau|x_1, \ldots, x_n) &= \frac{f(x_1, \ldots, x_n|\tau)f(\tau)}{f(x_1, \ldots, x_n)} \\
&= \frac{\left(\frac{\tau}{2\pi}\right)^{\frac{1}{2}n} e^{-\frac{1}{2}\tau n s^2} \frac{\beta^\alpha}{\Gamma(\alpha)}\tau^{\alpha-1}e^{-\beta\tau}}{\left(\frac{1}{2\pi}\right)^{\frac{1}{2}n} \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \tau^{\frac{1}{2}n}\tau^{\alpha-1}e^{-\frac{1}{2}\tau n s^2}e^{-\beta\tau}d\tau} \\
&= \frac{\tau^{\alpha+\frac{1}{2}n-1}e^{-\tau(\frac{1}{2}ns^2+\beta)}}{\int_0^\infty \tau^{\alpha+\frac{1}{2}n-1}e^{-\tau(\frac{1}{2}ns^2+\beta)}d\tau} \\
&= \frac{\tau^{\alpha+\frac{1}{2}n-1}e^{-\tau(\frac{1}{2}ns^2+\beta)}}{\frac{\Gamma(\alpha+\frac{1}{2}n)}{(\frac{1}{2}s^2n+\beta)^{\alpha+n/2}}} \\
&= \frac{\left(\frac{1}{2}s^2n+\beta\right)^{\alpha+\frac{1}{2}n}}{\Gamma(\alpha+\frac{1}{2}n)}\tau^{\alpha+\frac{1}{2}n-1}e^{-\tau(\frac{1}{2}ns^2+\beta)} \\
&\sim \mathcal{G}\left(\alpha+n/2, ns^2/2+\beta\right),
\end{aligned}
$$

as was to prove. $\qquad\square$

F.4 Posterior distribution of an Exponential variable with Gamma distributed parameter

Theorem F.6 *Suppose a random variable T is distributed $\mathcal{E}(\theta)$ Suppose also θ is distributed according to a Gamma distribution $\mathcal{G}(\alpha, \beta)$ with density (F.1). Then the posterior density $f(\theta|t)$ is distributed $\mathcal{G}(\alpha+1, \beta+t)$.*

Proof: the posterior density equals

$$
\begin{aligned}
f(\theta|t) &= \frac{f(\theta, t)}{f(t)} \\
&= \frac{\theta e^{-\theta t} \frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}}{\frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \theta^\alpha e^{-\theta(t+\beta)}} \\
&= \frac{\theta^\alpha e^{-\theta(t+\beta)}}{\frac{\Gamma(\alpha+1)}{(t+\beta)^{\alpha+1}}} \\
&= \frac{(t+\beta)^{\alpha+1}}{\Gamma(\alpha+1)}\theta^\alpha e^{-\theta(t+\beta)} \\
&\sim \mathcal{G}(\alpha+1, \beta+t)
\end{aligned}
$$

as was to prove. $\qquad\square$

Bibliography

Andersen P.K., Borgan Ø, Gill R.D. and Keiding N. (1993). *Statistical models based on counting processes* Springer-Verlag, New York.

Anderson T.W. (1984). *An introduction to multivariate statistical analysis* 2nd. edition, Wiley, New York.

Apostol T.M. (1967). *Calculus, Volume I* 2nd edition, Wiley, New York.

Baks K. (1995). Bayes endogenously stratified sampling. MSc thesis University of Groningen/Brown University.

Casella G. and George E.I. (1992). Explaining the gibbs sampler *The American statistician* **46** 167-174.

CBS (1991). *Sociaal-economisch panelonderzoek. Inhoud, opzet en organisatie* (in dutch) Internal report CBS.

Erdélyi A., Magnus W., Oberhettinger F., Tricomi F.G. (1953). *Higher transcendental functions, Volume I* California Institute of Technology, McGraw-Hill Book Company, New York.

Gelfand A.E., Hills S.E., Racine-Poon A. and Smith A.F.M. (1990). Illustration of bayesian inference in normal data models using gibbs sampling. *Journal of the American Statistical Association* **85** 972-985.

Gelfand A.E., Smith A.F.M. and Tai-Ming Lee (1992). Bayesian analysis of constrained parameter and truncated data problems using gibbs sampling. *Journal of the American Statistical Association* **87** 523-532.

Geman S. and Geman D. (1984). Stochastic relaxation. Gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6** 721-741.

Gorter D. (1991). *The relation between unemployment benefit, reservation wage and search duration* Internal report CBS.

Gorter D. and Hoogteijling E. (1990). *Duur van het zoeken naar werk* (in dutch) Internal report CBS.

Hendriks M.M.W.B. (1996). Nonlinear retention modeling in liquid chromatography. PhD thesis. University of Groningen.

Johnston J. (1984). *Econometric methods* 3rd. edition, McGraw-Hill Book Company, New York.

Jones S.R.G. (1988). The relationship between unemployment spells and reservation wages as a test of search theory. *Quarterly Journal of Economics* **103** 741-765.

Kalfleisch J.D. and Prentice R.L. (1980). *The statistical analysis of failure time data* Wiley, New York.

Kortram R.A., van Rooij A.C.M., Lenstra A.J. and Ridder G. (1995). Constructive identification of the mixed proportional hazards model. *Statistica Neerlandica* **49** 269-281.

Kroese, A.H. (1994). Distributional inference: a loss function approach. PhD thesis. University of Groningen.

Lancaster T. (1990). *The econometric analysis of transition data* Cambridge university press, Cambridge.

Lancaster T. and Chesher A. (1983). An econometric analysis of reservation wages *Econometrica* **51** 1661-1676.

Leunis W.P. and Altena J.W. (1996). Labour accounts in the Netherlands, 1987-1993. How to cope with fragmented macro data in official statistics. *International Statistical Review* **64** 1-22.

Mikosch T. (1994). *Empirical processes* Lecture notes, University of Groningen.

Mortensen D.T. (1986). Job search and labor market analysis, in: O. Ashenfelter and R. Layard (eds.), *Handbook of Labor economics, Volume II*, Elsevier Science Publishers BV, 1986

Neyman J. (1937). "Smooth" test for goodness of fit. *Skand. Aktuarietidkr.* **20** 150-199.

Odell P.L. and Feiveson A.H. (1966). A numerical procedure to generate a sample covariance matrix. *Journal of the American Statistical Association* **61** 198-203.

O'Hagan A. (1994). *Kendall's advanced theory of statistics, Volume 2B, Bayesian inference* Edward Arnold, London.

Rayner J.C.W. and Best D.J. (1989). *Smooth tests of goodness of fit*, Oxford university press, New York.

Ryshik I.M. and Gradstein I.S. (1963). *Tables of series, products and integrals*, Veb Deutscher Verlag der wissenschaften, Berlin.

Tanner M.A. (1993). *Tools for statistical inference* 2nd edition, Springer-Verlag, New York.

Smith A.F.M. and Roberts G.O. (1993). Bayesian computation via the gibbs sampler and related markov chain monte carlo methods. *Journal of the Royal Statistical Society B* **55** 3-23.

Stewart J. (1991). *Econometrics* Philip Allan, New York.

Zellner A. (1971). *An introduction to Bayesian inference in econometrics* Wiley, New York.