Chapter 14

Maximum Likelihood Estimation

**Exercises**

1. The density of the maximum is

*n*[*z*/θ]*n*-1(1/θ), 0 < *z* < θ.

Therefore, the expected value is *E*[*z*] = *zndz* = [θ*n*+1/(*n*+1)][*n*/θ*n*] = *n*θ/(*n*+1). The variance is found likewise, *E*[*z*2] = *z*2*n*(*z*/*n*)*n*-1(1/θ)*dz* = *n*θ2/(*n*+2) so Var[*z*] = *E*[*z*2] ‑ (*E*[*z*])2 = *n*θ2/[(*n* + 1)2(*n*+2)]. Using mean squared convergence we see that *E*[*z*] = θ and Var[*z*] = 0, so that plim *z* = θ.

2. The log‑likelihood is ln*L* = ‑*n*lnθ ‑ (1/θ). The maximum likelihood estimator is obtained as the solution to ∂ln*L*/∂θ = ‑*n*/θ + (1/θ2)  = 0, or  = . The asymptotic variance of the MLE is {‑*E*[∂2ln*L*/∂θ2]}-1 = {‑*E*[*n*/θ2 ‑ (2/θ3) ]}-1. To find the expected value of this random variable, we need *E*[*x*i] = θ. Therefore, the asymptotic variance is θ2/*n*. The asymptotic distribution is normal with mean θ and this variance.

3. The log‑likelihood is ln*L* = *n*lnθ ‑ (β+θ) + lnβ + ‑ 

The first and second derivatives are ∂ln*L*/∂θ = *n*/θ‑

∂ln*L*/∂β = ‑ + /β

∂2ln*L*/∂θ2 = ‑*n*/θ2

∂2ln*L*/∂β2 = ‑ /β2

∂2ln*L*/∂β∂θ = 0.

Therefore, the maximum likelihood estimators are  = 1/ and  =  and the asymptotic covariance matrix is the inverse of . In order to complete the derivation, we will require the expected value of = *nE*[*xi*]. In order to obtain *E*[*xi*], it is necessary to obtain the marginal distribution of *xi*, which is f(x) = =  This is β*x*(θ/*x*!) times a gamma integral. This is *f*(*x*) = β*x*(θ/*x*!)[Γ(*x*+1)]/(β+θ)*x*+1. But, Γ(*x*+1) = *x*!, so the expression reduces to

*f*(*x*) = [θ/(β+θ)][β/(β+θ)]*x*.

Thus, *x* has a geometric distribution with parameter π = θ/(β+θ). (This is the distribution of the number of tries until the first success of independent trials each with success probability 1‑π. Finally, we require the expected value of *xi*, which is *E*[*x*] = [θ/(β+θ)] *x*[β/(β+θ)]*x*= β/θ. Then, the required asymptotic covariance matrix is .

The maximum likelihood estimator of θ/(β+θ) is is

= (1/)/[/ + 1/] = 1/(1 + ).

Its asymptotic variance is obtained using the variance of a nonlinear function

*V* = [β/(β+θ)]2(θ2/*n*) + [-θ/(β+θ)]2(βθ/*n*) = βθ2/[*n*(β+θ)3].

The asymptotic variance could also be obtained as [-1/(1 + *E*[*x*])2]2Asy.Var[].)

For part (c), we just note that γ = θ/(β+θ). For a sample of observations on *x*, the log‑likelihood would be ln*L* = *n*lnγ + ln(1‑γ)

∂ln*L*/dγ = n/γ ‑ /(1‑γ).

A solution is obtained by first noting that at the solution, (1‑γ)/γ =  = 1/γ ‑ 1. The solution for γ is, thus,

= 1 / (1 +).Of course, this is what we found in part b., which makes sense.

For part (d) *f*(*y*|*x*) =  =  Cancelling terms and gathering the remaining like terms leaves *f*(*y*|*x*) =  so the density has the required form with λ = (β+θ). The integral is . This integral is a Gamma integral which equals Γ(*x*+1)/λ*x*+1, which is the reciprocal of the leading scalar, so the product is 1. The log‑likelihood function is

ln*L* = *n*lnλ ‑ λ + lnλ ‑ 

∂ln*L*/∂λ = (+ *n*)/λ ‑ .

∂2ln*L*/∂λ2 = ‑(+ *n*)/λ2.

Therefore, the maximum likelihood estimator of λ is (1 + )/ and the asymptotic variance, conditional on the *x*s is Asy.Var. = (λ2/*n*)/(1 +)

Part (e.) We can obtain *f*(*y*) by summing over *x* in the joint density. First, we write the joint density as . The sum is, therefore, . The sum is that of the probabilities for a Poisson distribution, so it equals 1. This produces the required result. The maximum likelihood estimator of θ and its asymptotic variance are derived from

ln*L* = *n*lnθ ‑ θ

∂ln*L*/∂θ = *n*/θ ‑ 

∂2ln*L*/∂θ2 = ‑*n*/θ2.

Therefore, the maximum likelihood estimator is 1/ and its asymptotic variance is θ2/*n*. Since we found *f*(*y*) by factoring *f*(*x*,*y*) into *f*(*y*)*f*(*x*|*y*) (apparently, given our result), the answer follows immediately. Just divide the expression used in part e. by *f*(*y*). This is a Poisson distribution with parameter β*y*. The log‑likelihood function and its first derivative are

ln*L* = ‑β + ln +  ‑ 

∂ln*L*/∂β = ‑ + /β,

from which it follows that .

4. The log‑likelihood and its two first derivatives are

log*L* = *n*logα + *n*logβ + (β‑1) ‑ α

∂log*L*/∂α = *n*/α ‑ 

∂log*L*/∂β = *n*/β + ‑ α

Since the first likelihood equation implies that at the maximum, = *n* /, one approach would be to scan over the range of β and compute the implied value of α. Two practical complications are the allowable range of β and the starting values to use for the search.

The second derivatives are

∂2ln*L*/∂α2 = ‑*n*/α2

∂2ln*L*/∂β2 = ‑*n*/β2 ‑ α

∂2ln*L*/∂α∂β = ‑.

If we had estimates in hand, the simplest way to estimate the expected values of the Hessian would be to evaluate the expressions above at the maximum likelihood estimates, then compute the negative inverse. First, since the expected value of ∂ln*L*/∂α is zero, it follows that E[*x*iβ] = 1/α. Now,

*E*[∂ln*L*/∂β] = *n*/β + *E*[] ‑ α*E*[]= 0

as well. Divide by *n*, and use the fact that every term in a sum has the same expectation to obtain

1/β + *E*[ln*xi*] ‑ *E*[(ln*x*i)*xi*β]/*E*[*xi*β] = 0.

Now, multiply through by *E*[*xi*β] to obtain *E*[*xi*β] = *E*[(ln*xi*)*xi*β] ‑ *E*[ln*x*i]*E*[*xi*β]

or 1/(αβ) = Cov[ln*x*i,*xi*β]. ~

5. As suggested in the previous problem, we can concentrate the log‑likelihood over α. From ∂log*L*/∂α = 0, we find that at the maximum, α = 1/[(1/*n*) ]. Thus, we scan over different values of β to seek the value which maximizes log*L* as given above, where we substitute this expression for each occurrence of α. Values of β and the log‑likelihood for a range of values of β are listed and shown in the figure below.

β log*L*



1

0.1 ‑62.386

0.2 ‑49.175

0.3 ‑41.381

0.4 ‑36.051

0.5 ‑32.122

0.6 ‑29.127

0.7 ‑26.829

0.8 ‑25.098

0.9 ‑23.866

1.0 ‑23.101

1.05 ‑22.891

1.06 ‑22.863

1.07 ‑22.841

1.08 ‑22.823

1.09 ‑22.809

1.10 ‑22.800

1.11 ‑22.796

1.12 ‑22.797

1.2 ‑22.984

1.3 ‑23.693

The maximum occurs at β = 1.11. The implied value of α is 1.179. The negative of the second derivatives matrix at these values and its inverse are  and .

The Wald statistic for the hypothesis that β = 1 is *W* = (1.11 ‑ 1)2/.041477 = .276. The critical value for a test of size .05 is 3.84, so we would not reject the hypothesis.

If β = 1, then  =  = 0.88496. The distribution specializes to the geometric distribution if β = 1, so the restricted log‑likelihood would be

log*Lr* = nlogα ‑ α = *n*(logα - 1) at the MLE.

log*Lr*at α = .88496 is ‑22.44435. The likelihood ratio statistic is ‑2logλ = 2(23.10068 ‑ 22.44435) = 1.3126.

Once again, this is a small value. To obtain the Lagrange multiplier statistic, we would compute



at the restricted estimates of α = .88496 and β = 1. Making the substitutions from above, at these values, we would have

∂log*L*/∂α = 0

∂log*L*/∂β = *n* + ‑  = 9.400342

∂2log*L*/∂α2 =  = ‑25.54955

∂2log*L*/∂β2 = ‑*n* ‑ = ‑30.79486

∂2log*L*/∂α∂β =  = ‑8.265.

The lower right element in the inverse matrix is .041477. The LM statistic is, therefore, (9.40032)2.041477 = 2.9095. This is also well under the critical value for the chi‑squared distribution, so the hypothesis is not rejected on the basis of any of the three tests.

6. a. The full log likelihood is logL = Σ log fyx(y,x|α,β).

b. By factoring the density, we obtain the equivalent logL = Σ[ log fy|x (y|x,α,β) + log fx (x|α)]

c. We can solve the first order conditions in each case. From the marginal distribution for x,

Σ ∂ log fx (x|α)/∂α = 0

provides a solution for α. From the joint distribution, factored into the conditional plus the marginal, we have

Σ[ ∂log fy|x (y|x,α,β)/∂α + ∂log fx (x|α)/∂α = 0

Σ[ ∂log fy|x (y|x,α,β)/∂β = 0

d. The asymptotic variance obtained from the first estimator would be the negative inverse of the expected second derivative, Asy.Var[a] = {[-E[Σ2∂ log fx (x|α)/∂α2]}-1. Denote this Aαα-1. Now, consider the second estimator for α and β jointly. The negative of the expected Hessian is shown below. Note that the Aαα from the marginal distribution appears there, as the marginal distribution appears in the factored joint distribution.



The asymptotic covariance matrix for the joint estimator is the inverse of this matrix. To compare this to the asymptotic variance for the marginal estimator of α, we need the upper left element of this matrix. Using the formula for the partitioned inverse, we find that this upper left element in the inverse is

[(Aαα+Bαα) - (BαβBββ-1Bβα)]-1 = [Aαα + (Bαα - BαβBββ-1Bβα)]-1

which is smaller than Aαα as long as the second term is positive.

e. (Unfortunately, this is an error in the text.) In the preceding expression, Bαβ is the cross derivative. Even if it is zero, the asymptotic variance from the joint estimator is still smaller, being [Aαα + Bαα]-1. This makes sense. If α appears in the conditional distribution, then there is additional information in the factored joint likelhood that is not in the marginal distribution, and this produces the smaller asymptotic variance.

7. The log likelihood for the Poisson model is

LogL = -nλ + logλΣi yi - Σi log yi!

The expected value of 1/n times this function with respect to the true distribution is

E[(1/n)logL] = -λ + logλ E0[] – E0 (1/n)Σi logyi!

The first expectation is λ0. The second expectation can be left implicit since it will not affect the solution for λ - it is a function of the true λ0. Maximizing this function with respect to λ produces the necessary condition

∂E0 (1/n)logL]/∂λ = -1 + λ0/λ = 0

which has solution λ = λ0 which was to be shown.

8. The log likelihood for a sample from the normal distribution is

LogL = -(n/2)log2π - (n/2)logσ2 – 1/(2σ2) Σi (yi - μ)2.

E0 [(1/n)logL] = -(1/2)log2π - (1/2)logσ2 – 1/(2σ2) E0[(1/n) Σi (yi - μ)2].

The expectation term equals E0[(yi - μ)2] = E0[(yi - μ0)2] + (μ0 - μ)2 = σ02 + (μ0 - μ)2 . Collecting terms,

E0 [(1/n)logL] = -(1/2)log2π - (1/2)logσ2 – 1/(2σ2)[ σ02 + (μ0 - μ)2]

To see where this is maximized, note first that the term (μ0 - μ)2 enters negatively as a quadratic, so the maximizing value of μ is obviously μ0. Since this term is then zero, we can ignore it, and look for the σ2 that maximizes -(1/2)log2π - (1/2)logσ2 – σ02/(2σ2). The –1/2 is irrelevant as is the leading constant, so we wish to minimize (after changing sign) logσ2 + σ02/σ2 with respect to σ2. Equating the first derivative to zero produces 1/σ2 = σ02/(σ2)2 or σ2 = σ02, which gives us the result.

9. The log likelihood for the classical normal regression model is

LogL = Σi -(1/2)[log2π + logσ2 + (1/σ2)(yi - xi′β)2]

If we reparameterize this in terms of η = 1/σ and δ = β/σ, then after a bit of manipulation,

LogL = Σi -(1/2)[log2π - logη2 + (ηyi - xi′δ)2]

The first order conditions for maximizing this with respect to η and δ are

∂logL/∂η = n/η - Σi yi (ηyi - xi′δ) = 0

∂logL/∂δ = Σi xi (ηyi - xi′δ) = 0

Solve the second equation for δ, which produces δ = η (X′X)-1X′y = η b. Insert this implicit solution into the first equation to produce n/η = Σi yi (ηyi - ηxi′b). By taking η outside the summation and multiplying the entire expression by η, we obtain n = η2 Σi yi (yi - xi′b) or η2 = n/[Σi yi (yi - xi′b)]. This is an analytic solution for η that is only in terms of the data – b is a sample statistic. Inserting the square root of this result into the solution for δ produces the second result we need. By pursuing this a bit further, you canshow that the solution for η2 is just n/e′e from the original least squares regression, and the solution for δ is just b times this solution for η. The second derivatives matrix is

∂2logL/∂η2 = -n/η2 - Σiyi2

∂2logL/∂δ ∂δ′ = -Σi xixi′

∂2logL/∂δ ∂η = Σi xiyi.

We’ll obtain the expectations conditioned on X. E[yi|xi] is xi′β from the original model, which equals xi′δ/η. E[yi2|xi] = 1/η2 (δ′xi)2 + 1/η2. (The cross term has expectation zero.) Summing over observations and collecting terms, we have, conditioned on X,

E[∂2logL/∂η2|X] = -2n/η2 - (1/η2)δ′X′Xδ

E[∂2logL/∂δ ∂δ′|X] = -X′X

E[∂2logL/∂δ ∂η|X] = (1/η)X′Xδ

The negative inverse of the matrix of expected second derivatives is



(The off diagonal term does not vanish here as it does in the original parameterization.)

10**.**  The first derivatives of the log likelihood function are ∂logL/∂μ = -(1/2σ2) Σi -2(**y**i - **μ**). Equating this to zero produces the vector of means for the estimator of **μ**. The first derivative with respect to σ2 is

∂logL/∂σ2 = -nM/(2σ2) + 1/(2σ4)Σi (**y**i - **μ**)**′**(**y**i - **μ**). Each term in the sum is Σm (yim - μm)2. We already deduced that the estimators of μm are the sample means. Inserting these in the solution for σ2 and solving the likelihood equation produces the solution given in the problem. The second derivatives of the log likelihood are

∂2logL/∂**μ**∂**μ′** = (1/σ2)Σ i -**I**

∂2logL/∂**μ**∂σ2 = (1/2σ4) Σi -2(**y**i - **μ**)

∂2logL/∂σ2∂σ2 = nM/(2σ4) - 1/σ6 Σi (**y**i - **μ**)**′**(**y**i - **μ**)

The expected value of the first term is (-n/σ2)**I**. The second term has expectation zero. Each term in the summation in the third term has expectation Mσ2, so the summation has expected value nMσ2. Adding gives the expectation for the third term of -nM/(2σ4). Assembling these in a block diagonal matrix, then taking the negative inverse produces the result given earlier.

For the Wald test, the restriction is

H0: **μ** - μ0**i** = **0**.

The unrestricted estimator of μ is . The variance of  is given above, so the Wald statistic is simply

( - μ0**i** )′ Var[( - μ0**i** )]-1( - μ0**i** ). Inserting the covariance matrix given above produces the suggested statistic.

11. The asymptotic variance of the MLE is, in fact, equal to the Cramer-Rao Lower Bound for the variance of a consistent, asymptotically normally distributed estimator, so this completes the argument.

In example 4.7, we proposed a regression with a gamma distributed disturbance,

*yi* = α + **x***i***′*β*** + *εi*

where,

*f*(*εi*) = [λ*P*/Γ(*P*)] *εiP*-1 exp(-λ*εi*), *εi* > 0, λ > 0, *P* > 2.

(The fact that *εi* is nonnegative will shift the constant term, as shown in Example 4.7. The need for the restriction on *P* will emerge shortly.) It will be convenient to assume the regressors are measured in deviations from their means, so Σ*i***x***i* = **0**. The OLS estimator of ***β*** remains unbiased and consistent in this model, with variance

Var[**b**|**X**] = σ2(**X′X**)-1

where σ2 = Var[*εi*|**X**] = *P*/λ2. [You can show this by using gamma integrals to verify that *E*[*εi*|**X**] = *P*/λ and E[*εi*2|**X**] = *P*(*P*+1)/λ2. See B-39 and (E-1) in Section E2.3. A useful device for obtaining the variance is Γ(*P*) = (*P*-1)Γ(*P*-1).] We will now show that in this model, there is a more efficient consistent estimator of ***β***. (As we saw in Example 4.7, the constant term in this regression will be biased because *E*[*εi*|**X**] = *P*/λ; *a* estimates α+*P*/λ. In what follows, we will focus on the slope estimators.

The log likelihood function is

Ln *L* = 

The likelihood equations are

∂ ln*L*/∂α = Σ*i* [-(*P*-1)/*εi* + λ] = 0,

∂ ln*L*/∂**β** = Σ*i* [-(*P*-1)/ε*i* + λ]**x***i* = **0**,

∂ ln*L*/∂λ = Σ*i* [*P*/λ - ε*i*] = 0,

∂ ln*L*/∂*P* = Σ*i* [lnλ - ψ(*P*) - *εi*] = 0.

The function ψ(*P*) = dlnΓ(*P*)/d*P* is defined in Section E2.3.) To show that these expressions have expectation zero, we use the gamma integral once again to show that *E*[1/*εi*] = λ/(*P*-1). We used the result *E*[ln*εi*] = ψ(*P*)-λ in Example 13.5. So show that *E*[∂ln*L*/∂**β**] = **0**, we only require *E*[1/*εi*] = λ/(*P*-1) because **x***i* and *εi* are independent. The second derivatives and their expectations are found as follows: Using the gamma integral once again, we find *E*[1/*εi*2] = λ2/[(*P*-1)(*P*-2)]. And, recall that Σ*i***x***i* = **0**. Thus, conditioned on **X**, we have

-*E*[∂2lnL/∂α2] = *E*[Σ*i* (*P*-1)(1/*εi*2)] = *n*λ2/(*P*-2),

-*E*[∂2lnL/∂α∂***β***] = *E*[Σ*i* (*P*-1)(1/*εi*2)**x***i*] = **0,**

-*E*[∂2lnL/∂α∂λ] = *E*[Σ*i* (-1)] = -*n*,

-*E*[∂2lnL/∂α∂*P*] = *E*[Σ*i* (1/*εi*)] = *n*λ/(*P*-1),

-*E*[∂2lnL/∂***β***∂***β*′**] = *E*[Σ*i* (*P*-1)(1/*εi*2)**x***i***x***i***′**] = Σi [λ2/(*P*-2)]**x***i***x***i***′ =** [λ2/(*P*-2)](**X′X**),

-*E*[∂2lnL/∂λ∂***β***] = *E*[Σ*i* (-1)**x***i*] = **0**,

-*E*[∂2lnL/∂*P*∂***β***] = *E*[Σ*i* (1/*εi*)**x***i*] = **0**,

-*E*[∂2lnL/∂λ2] = *E*[Σ*i* (*P*/λ2)] = *nP*/λ2,

-*E*[∂2lnL/∂λ∂*P*] = *E*[Σ*i* (1/λ)] = *n*/λ,

-*E*[∂2lnL/∂P2] = *E*[Σ*i* ψ′(*P*)] = *n*ψ**′**(*P*).

Since the expectations of the cross partials witth respect to ***β*** and the other parameters are all zero, it follows that the asymptotic covariance matrix for the MLE of ***β*** is simply

Asy.Var[] = {-*E*[∂2lnL/∂***β***∂**β′**]}-1 = [(*P*-2)/λ2](**X′X**)-1.

Recall, the asymptotic covariance matrix of the ordinary least squares estimator is

Asy.Var[**b**] = [*P*/λ2](**X′X**)-1.

(Note that the MLE is ill defined if *P* is less than 2.) Thus, the ratio of the variance of the MLE of any element of ***β*** to that of the corresponding element of **b** is (*P*-2)/*P* which is the result claimed in Example 4.7.

**Applications**

1. a. For both probabilities, the symmetry implies that 1 – *F*(*t*) = *F*(-*t*). In either model, then,

Prob(*y*=1) = *F*(t) and Prob(*y*=0) = 1 – *F*(t) = *F*(-*t*).

These are combined in Prob(*Y*=*y*) = F[(2*yi*-1)*ti*] where *ti* = **xi′β**. Therefore,

ln *L* = Σi ln *F*[(2*y*i-1)**x**i′**β**]

b. ∂lnL/∂**β** = = **0**

where *f*[(2*yi*-1)**x**i**′β**] is the density function. For the logit model, *f* = *F*(1-*F*). So, for the logit model,

∂lnL/∂**β** = = **0**

Evaluating this expression for *yi* = 0, we get simply –*F*(**x**i**′β**)**x**i. When *yi* = 1, the term is

[1- *F*(**x**i**′β**)]**x**i. It follows that both cases are [*yi* - *F*(**x**i**′β**)]**x**i, so the likelihood equations for the logit model are

∂lnL/∂**β** = = **0.**

For the probit model, *F*[(2*y*i-1)**x**i′**β**] = Φ[(2*y*i-1)**x**i′**β**] and f[(2*y*i-1)**x**i′**β**] = φ[(2*y*i-1)**x**i′**β**], which does not simplify further, save for that the term 2yi inside may be dropped since φ(t) = φ(-t). Therefore,

∂lnL/∂**β** = = **0**

c. For the logit model, the result is very simple.

∂2lnL/∂**β**∂**β**′= **.**

For the probit model, the result is more complicated. We will use the result that

dφ(t)/dt = -tφ(t).

It follows, then, that d[φ(t)/Φ(t)]/dt = [-φ(t)/Φ(t)][t + φ(t)/Φ(t)]. Using this result directly, it follows that

∂2lnL/∂**β**∂**β**′= = **0**

This actually simplifies somewhat because (2yi-1)2 = 1 for both values of yi and = 

d. Denote by **H** the actual second derivatives matrix derived in the previous part. Then, Newton’s method is



where the terms on the right hand side indicate first and second derivatives evaluated at the “previous” estimate of β.

e. The method of scoring uses the expected Hessian instead of the actual Hessian in the iterations. The methods are the same for the logit model, since the Hessian does not involve yi. The methods are different for the probit model, since the expected Hessian does not equal the actual one. For the logit model

-[*E*(**H**)]-1 = 

For the probit model, we need first to obtain the expected value. Do obtain this, we take the expected value, with Prob(y=0) = 1 - Φ and Prob(y=1) = Φ. The expected value of the ith term in the negative hessian is the expected value of the term,



This is







?====================================================

? Application 14.1 Part f.

?====================================================

Namelist ; x = one,age,educ,hsat,female,married $

LOGIT ; Lhs = Doctor ; Rhs = X $

Calc ; L1 = logl $

+---------------------------------------------+

| Binary Logit Model for Binary Choice |

| Dependent variable DOCTOR |

| Number of observations 27326 |

| Log likelihood function -16405.94 |

| Number of parameters 6 |

| Info. Criterion: AIC = 1.20120 |

| Info. Criterion: BIC = 1.20300 |

| Restricted log likelihood -18019.55 |

+---------------------------------------------+

+--------+--------------+----------------+--------+--------+----------+

|Variable| Coefficient | Standard Error |b/St.Er.|P[|Z|>z]| Mean of X|

+--------+--------------+----------------+--------+--------+----------+

---------+Characteristics in numerator of Prob[Y = 1]

Constant| 1.82207669 .10763712 16.928 .0000

AGE | .01235692 .00124643 9.914 .0000 43.5256898

EDUC | -.00569371 .00578743 -.984 .3252 11.3206310

HSAT | -.29276744 .00686076 -42.673 .0000 6.78542607

FEMALE | .58376753 .02717992 21.478 .0000 .47877479

MARRIED | .03550015 .03173886 1.119 .2634 .75861817

g.

Matr ; bw = b(5:6) ; vw = varb(5:6,5:6) $

Matrix ; list ; WaldStat = bw'<vw>bw $

Calc ; list ; ctb(.95,2) $

LOGIT ; Lhs = Doctor ; Rhs = One,age,educ,hsat $

Calc ; L0 = logl $

Calc ; List ; LRStat = 2\*(l1-l0) $

Matrix WALDSTAT has 1 rows and 1 columns.

1

+--------------

1| 461.43784

**--> Calc ; list ; ctb(.95,2) $**

+------------------------------------+

| Listed Calculator Results |

+------------------------------------+

Result = 5.991465

**--> Calc ; L0 = logl $**

**--> Calc ; List ; LRStat = 2\*(l1-l0) $**

+------------------------------------+

| Listed Calculator Results |

+------------------------------------+

LRSTAT = 467.336374

Logit ; Lhs = Doctor ; Rhs = X ; Start = b,0,0 ; Maxit = 0 $

+---------------------------------------------+

| Binary Logit Model for Binary Choice |

| Maximum Likelihood Estimates |

| Model estimated: May 17, 2007 at 11:49:42PM.|

| Dependent variable DOCTOR |

| Weighting variable None |

| Number of observations 27326 |

| Iterations completed 1 |

| LM Stat. at start values 466.0288 |

| LM statistic kept as scalar LMSTAT |

| Log likelihood function -16639.61 |

| Number of parameters 6 |

| Info. Criterion: AIC = 1.21830 |

| Finite Sample: AIC = 1.21830 |

| Info. Criterion: BIC = 1.22010 |

| Info. Criterion:HQIC = 1.21888 |

| Restricted log likelihood -18019.55 |

| McFadden Pseudo R-squared .0765802 |

| Chi squared 2759.883 |

| Degrees of freedom 5 |

| Prob[ChiSqd > value] = .0000000 |

| Hosmer-Lemeshow chi-squared = 23.44388 |

| P-value= .00284 with deg.fr. = 8 |

+---------------------------------------------+

h. The restricted log likelihood given with the initial results equals -18019.55. This is the log likelihood for a model that contains only a constant term. The log likelihood for the model is

-16405.94. Twice the difference is about 3,200, which vastly exceeds the critical chi squared with 5 degrees of freedom. The hypothesis would be rejected.

Chapter 15

**Simulation Based Estimation and**

**Inference and Random Parameter Models**

**Exercises**

1. Exponential: The pdf is f(x) = θexp(-θx). The CDF is



We would draw observations from the U(0,1) population, say Fi, and equate these to F(xi). Inverting the function, we find that 1-Fi = exp(-θxi), or –(1/θ)ln(1-Fi) = xi.

2. Weibull. If the survival function is S(x) = λpexp[-(λx)p], then we may equate random draws from the uniform distribution, Si to this function (a draw of Si is the same as a draw of Fi = 1-Si). Solving for xi, we find

lnSi = ln(λp) – (λx)p, so xi = (1/λ)[ln(λp) – lnSi]1/p.

3. The derivative of the simulated sum of squares is



To estimate the asymptotic covariance matrix, we rely can rely on the results for the nonlinear least squares estimator. The gradient is of the form ΣiΣt -2eit xit0. The approach taken in Theorem 7.2 would be as follows:

. The counterpart to (**X′X**)-1 is



**Application**

?================================================================

? Application 15.1. Monte Carlo Simulation

?================================================================

? Set seed of RNG for replicability

Calc ; Ran(123579) $

? Sample size is 50. Generate x(i) and z(i) held fixed

Sample ; 1 - 50 $

Create ; xi = rnn(0,1) ; zi = rnn(0,1) $

Namelist ; X = one,xi,zi ; X0 = one,xi $

? Moment Matrices

Matrix ; XXinv = <X'X> ; X0X0inv = <X0'X0> $

Matrix ; Waldi = init(1000,1,0) $

Matrix ; LMi = init(1000,1,0) $

?\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

? Procedure studies the LM statistic

?\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Proc = LM (c) $

? Three kinds of disturbances

Create ?; Eps = Rnt(5) ? Nonnormal distribution

; vi=exp(.2\*xi) ; eps = vi\*rnn(0,1) ? Heteroscedasticity

?;eps= Rnn(0,1) ? Standard normal distribution

; y = 0 + xi + c\*zi +eps $

Matrix ; b0 = X0X0inv\*X0'y $

Create ; e0 = y - X0'b0 $

Matrix ; g = X'e0 $

Calc ; lmstat = qfr(g,xxinv)/(e0'e0/n) ; i = i + 1 $

Matrix ; Lmi (i) = lmstat $

EndProc $

Calc ; i = 0 ; gamma = -1 $

Exec ; Proc=LM(gamma) ; n = 1000 $

samp;1-1000$

create;LMv=lmi $

create;reject=lmv>3.84$

Calc ; List ; Type1 = xbr(reject) ; pwr = 1-Type1 $

?\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

? Procedure studies the Wald statistic

?\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Proc = Wald(c) $

Create ; if(type=1)Eps = Rnn(0,1) ? Standard normal distribution

; if(type=2)vi=exp(.2\*xi) ? eps = vi\*rnn(0,1) ? Heteroscedasticity

; if(type=3)eps= Rnt(5) ? Nonnormal distribution

; y = 0 + xi + c\*zi +eps $

Matrix ; b0=XXinv\*X'y $

Create ; e0=y-X'b0$

Calc ; ss0 = e0'e0/(47)

; v0 = ss0\*xxinv(3,3)

; wald0=(b0(3))^2/v0

; i=i+1 $

Matrix ; Waldi(i)=Wald0 $

EndProc $

? Set the values for the simulation

Calc ; i = 0 ; gamma = 0 ; type=1 $

Sample ; 1-50 $

Exec ; Proc=Wald(gamma) ; n = 1000 $

samp;1-1000$

create;Waldv=Waldi $

create;reject=Waldv > 3.84$

Calc ; List ; Type1 = xbr(reject) ; pwr = 1-Type1 $

To carry out the simulation, execute the procedure for different values of “gamma” and “type.” Summarize the results with a table or plot of the rejection probabilities as a function of gamma.

3. We will need a bivariate sample on x and y to compute the random variable, then average the draws on it. The precise method of using a Gibbs sampler to draw this bivariate sample is shown in Example 18.5. Once the bivariate sample of (x,y) is drawn, a large number of observations on [x2exp(y)+y2exp(x)] is computed and averaged. As noted there, the Gibbs sampler is not much of a simplification for this particular problem. It is simple to draw a sample dircectly from a bivariate normal distribution. Here is a program that does the simulation and plots the estimate of the function

Calc ; Ran(12345) $

Sample ; 1-1000$

Create ; xf=rnn(0,1) ; yfb=rnn(0,1) $

Matrix ; corr=init(100,1,0) ; function=corr $

Calc ; i=0 $

Proc

Calc ; i=i+1 $

Matrix ; corr(i)=ro $

Matrix ; c=[1/ro,1] ; c=chol(c) $

Create ; yf = c(2,1)\*xf + c(2,2)\*yfb $

Create ; fr=xf^2\*exp(yf)+yf^2\*exp(xf) $

Calc ; ef = xbr(fr) ; ro=ro+.02 $

Matrix ; function(i)=ef $

Endproc $

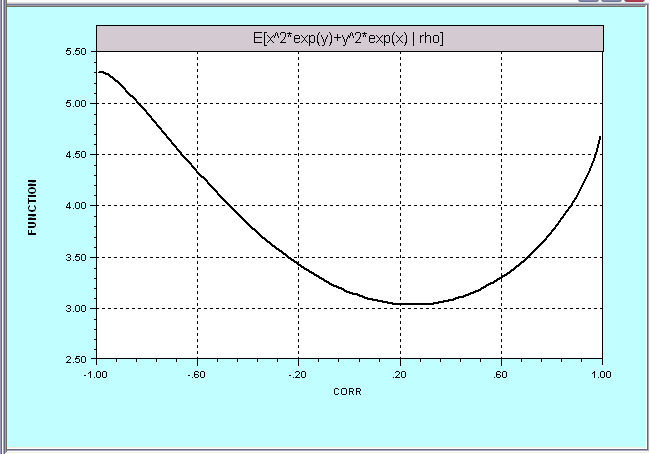
Calc ; ro=-.99 $

Execute; n=100 $

Mplot ; Lhs = corr ; Rhs = Function ; Fill

; Grid ; Endpoints = -1,1

; Title=E[x^2\*exp(y)+y^2\*exp(x) | rho] $



Chapter 16

**Bayesian Estimation and Inference**

**Exercise**

a. The likelihood function is

L(**y**|λ) = 

b. The posterior is

.

The product of factorials will fall out. This leaves



where we have used the gamma integral at the last step. The posterior defines a two parameter gamma distribution, G(n,).

c. The estimator of λ is the mean of the posterior. There is no need to do the integration. This falls simply out of the posterior density, E[λ|**y**] = /*n* = .

d. The posterior variance also drops out simply; it is /*n*2 = /*n*.

**Application**

a. *p*(*Fi*|*Ki*,θ) =  so the log likelihood function is



The MLE is obtained by setting ∂lnL(θ|y)/∂θ = Σi [Fi/θ - (Ki-Fi)/(1-θ)] = 0. Multiply both sides by θ(1-θ) to obtain

Σi [(1-θ)Fi - θ (Ki-Fi)] = 0

A line of algebra reveals that the solution is θ = (ΣiFi)/(ΣiKi) = 0.651596.

b. The posterior density is 

This simplifies considerably. The combinatorials and gamma functions fall out, leaving



The denominator is a beta integral, so the posterior density is



The denominator simplifies slightly;



c-e. The posterior distribution is a beta distribution with parameters a\*=(a+ΣiFi) and b\*=[b+Σi(Ki-Fi)].

The mean of this beta random variable is a\*/(a\*+b\*) = (a+ΣiFi)/(a+b+ΣiKi). In the data, Σi = 49 and ΣiKi = 75. For the values given, the posterior means are

(a=1,b=1): Result = .647668

(a=2,b=2): Result = .643939

(a=1,b=2): Result = .639386