15-19

likelihood estimator in the presence of fixed effects. The literature appears to take as given the qualitative wisdom of Hsiao and Abrevaya, that the FE/MLE is inconsistent when ~~with~~ T is small and fixed. (The implication that the severity of the inconsistency declines as T increases makes sense, but, again, remains to be shown analytically.)

The result for the two-period binary logit model is a standard result for discrete choice estimation. Several authors, all using Monte Carlo methods have pursued the result for the logit model for larger values of T. [See, for example, Katz (2001).] Greene (2004) analyzed the incidental parameters problem for other discrete choice models using Monte Carlo methods. We will examine part of that study.

The current studies are preceded by a small study in Heckman (1981) which examined the behavior of the fixed effects MLE in the following experiment:

$$z_{it} = 0.1t + 0.5z_{i,t-1} + u_{it}, z_{i0} = 5 + 10.0U_{i0},$$

lower case

lower case

$$U_{it} \sim U[-0.5, 0.5], i = 1, \ldots, 100, t = 0, \ldots, 8,$$

$$Y_{it} = \sigma_\tau \tau_i + \beta z_{it} + \varepsilon_{it}, \tau_i \sim N[0, 1], \varepsilon_{it} \sim N[0, 1],$$

$$y_{it} = 1 \text{ if } Y_{it} > 0, 0 \text{ otherwise.}$$

Heckman attempted to learn something about the behavior of the MLE for the probit model with $T = 8$. He used values of $\beta = -1.0, -0.1$, and 1.0 and $\sigma_\tau = 0.5, 1.0$, and 3.0. The mean values of the maximum likelihood estimates of β for the 9 cases are as follows: nine

	$\beta = -1.0$	$\beta = -0.1$	$\beta = 1.0$
$\sigma_\tau = 0.5$	−0.96	−0.10	0.93
$\sigma_\tau = 1.0$	−0.95	−0.09	0.91
$\sigma_\tau = 3.0$	−0.96	−0.10	0.90.

The findings here disagree with the received wisdom. Where there appears to be a bias (that is, excluding the center column), it seems to be quite small, and toward, not away from zero.

The Heckman study used a very small sample and, moreover, analyzed the fixed effects estimator in a random effects model (note that τ_i is independent of z_{it}). Greene (2004a), using the same parameter values, number of replications, and sample designs found persistent biases away from zero on the order of 15–20 percent. Numerous authors have extended the logit result for $T = 2$ with larger values of T, and likewise persistently found biases, away from zero, that diminish with increases in T. Greene (2004a) redid the experiment for the logit model, then replicated it for the probit and ordered probit models. The experiment is designed as follows: All models are based on the same index function and

$$w_{it} = \alpha_i + \beta x_{it} + \delta d_{it}, \quad \text{where } \beta = \delta = 1,$$

$$x_{it} \sim N[0, 1], d_{it} = \mathbf{1}[x_{it} + h_{it} > 0], \quad \text{where } h_{it} \sim N[0, 1],$$

$$\alpha_i = \sqrt{T}\bar{x}_i + \alpha_i, \alpha_i \sim N[0, 1]. \qquad v_i, v_i$$

The regressors d_{it} and x_{it} are constructed to be correlated. The random term h_{it} is used to produce independent variation in d_{it}. There is, however, no within group correlation in x_{it} or d_{it} built into the data generator. (Other experiments suggested that the marginal distribution of x_{it} mattered little to the outcome of the experiment.) The correlations between the variables are approximately 0.7 between x_{it} and d_{it}, 0.4 between α_i and x_{it}, and 0.2 between α_i and d_{it}. The individual effect is produced from independent

15-20

TABLE ~~17.2~~ 15.5 Means of Empirical Sampling Distributions, $N = 1{,}000$ Individuals Based on 200 Replications

	T = 2		T = 3		T = 5		T = 8		T = 10		T = 20	
	β	δ	β	δ	β	δ	β	δ	β	δ	β	δ
Logit Coeff	2.020	2.027	1.698	1.668	1.379	1.323	1.217	1.156	1.161	1.135	1.069	1.062
Logit M.E.[a]	1.676	1.660	1.523	1.477	1.319	1.254	1.191	1.128	1.140	1.111	1.034	1.052
Probit Coeff	2.083	1.938	1.821	1.777	1.589	1.407	1.328	1.243	1.247	1.169	1.108	1.068
Probit M.E.[a]	1.474	1.388	1.392	1.354	1.406	1.231	1.241	1.152	1.190	1.110	1.088	1.047
Ord. Probit	2.328	2.605	1.592	1.806	1.305	1.415	1.166	1.220	1.131	1.158	1.058	1.068

[a] Average ratio of estimated marginal effect to true marginal effect.

variation, ~~α_i~~ v_i as well as the group mean of x_{it}. The latter is scaled by $\sqrt{T}$ to maintain the unit variances of the two parts—without the scaling, the covariance between α_i and x_{it} falls to zero as T increases and $\bar{x}_i$ converges to its mean of zero). Thus, the data generator for the index function satisfies the assumptions of the fixed effects model. The sample used for the results below contains $n = 1{,}000$ individuals. The data-generating processes for the discrete dependent variables are as follows:

probit: $y_{it} = \mathbf{1}[w_{it} + \varepsilon_{it} > 0],\ \varepsilon_{it} \sim N[0, 1],$

ordered probit: $y_{it} = \mathbf{1}[w_{it} + \varepsilon_{it} > 0] + \mathbf{1}[w_{it} + \varepsilon_{it} > 3],\ \varepsilon_{it} \sim N[0, 1],$

logit: $y_{it} = \mathbf{1}[w_{it} + v_{it} > 0],\ v_{it} = \log[u_{it}/(1 - u_{it})],\ u_{it} \sim U[0, 1].$

(The three discrete dependent variables are described in Chapter ~~23~~.) 17

TB 15.5 ~~15.4~~ 15.5 Table ~~17.2~~ reports the results of computing the MLE with 200 replications. Models were fit with $T = 2, 3, 5, 8, 10$, and 20. (Note that this includes Heckman's experiment.) Each model specification and group size (T) is fit 200 times with random draws for ε_{it} or u_{it}. The data on the regressors were drawn at the beginning of each experiment (that is, for each T) and held constant for the replications. The table contains the average estimate of the coefficient and, for the binary choice models, the partial effects. The value at the extreme left corresponds to the received result, the 100 percent bias in the $T = 2$ case. The remaining values show, as intuition would suggest, that the bias decreases with increasing T. The benchmark case of $T = 8$, appears to be less benign than Heckman's results suggested. One encouraging finding for the model builder is that the biases in the estimated marginal effects appears to be somewhat less than for the coefficients. [Greene (2004b) extends this analysis to some other models, including the tobit and truncated regression models discussed in Chapter 24. The results there suggest that the conventional wisdom for the tobit model may not be correct—the incidental parameters problem seems to appear in the estimator of σ^2 in the tobit model, not in the estimators of the slopes. This is consistent with the linear regression model, but not with the binary choice models.]

delete [

18

delete]

~~17.5 SIMULATION-BASED ESTIMATION~~

~~The technique of maximum simulated likelihood (MSL) is essentially a classical sampling theory counterpart to the hierarchical Bayesian estimator considered in Chapter 18. Since the celebrated paper of Berry, Levinsohn, and Pakes (1995), and a related~~

15.6 SIMULATION BASED ESTIMATION

Sections 15.3 – 15.5 developed a set of tools for inference about model parameters using simulation methods. This section will describe methods for using simulation as part of the estimation process. The modeling framework arises when integrals that cannot be computed directly appear in the estimation criterion function (sum of squares, log likelihood, and so on). To illustrate, and begin the development, in Section 15.6.1, we will construct a nonlinear model with random effects. Section 15.6.2 will describe how simulation is used to evaluate integrals for maximum likelihood estimation. Section 15.6.3 will develop an application, the random effects regression model.

15.6.1 Random Effects in a Nonlinear Model

In Example 11.16, we considered a nonlinear regression model for the number of doctor visits in the German Socioeconomic Panel. The basic form of the nonlinear regression model is

$$E[y_{it}|\mathbf{x}_{it}] = \exp(\mathbf{x}_{it}'\boldsymbol{\beta}),\ t = 1,\ldots,T_i,\ i = 1,\ldots,n.$$

In order to accommodate unobserved heterogeneity in the panel data, we extended the model to include a random effect,

$$E[y_{it}|\mathbf{x}_{it},u_i] = \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + u_i), \tag{15-9}$$

where u_i is an unobserved random effect with zero mean and constant variance, possibly normally distributed – we will turn to that shortly. We will now go a step further and specify a particular probability distribution for y_{it}. Since it is a count, the Poisson regression model would be a natural choice,

$$p(y_{it}\,|\,\mathbf{x}_{it},u_i) = \frac{\exp(-\mu_{it})\mu_{it}^{y_{it}}}{y_{it}!},\ \mu_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + u_i). \tag{15-10}$$

them/ Conditioned on $\mathbf{x}_{it}$, and u_i, the T_i observations for individual i are independent. That is, by conditioning on u_i, we treat it as data, the same as $\mathbf{x}_{it}$. Thus, the T_i observations are independent when they are conditioned on $\mathbf{x}_{it}$ and u_i. The joint density for the T_i observations for individual i is the product,

$$p(y_{i1},y_{i2},\ldots,y_{i,T_i}\,|\,\mathbf{X}_i,u_i) = \prod_{t=1}^{T_i}\frac{\exp(-\mu_{it})\mu_{it}^{y_{it}}}{y_{it}!},\ \mu_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + u_i),\ t = 1,\ldots,T_i. \tag{15-11}$$

In principle at this point, the log likelihood function to be maximized would be

$$\ln L = \sum_{i=1}^{n}\ln\left[\prod_{t=1}^{T_i}\frac{\exp(-\mu_{it})\mu_{it}^{y_{it}}}{y_{it}!}\right],\ \mu_{it} = \exp(\mathbf{x}_{it}'\boldsymbol{\beta} + u_i). \tag{15-12}$$

But, it is not possible to maximize this log likelihood because the unobserved u_i, $i = 1,\ldots,n$, appears in it. The joint distribution of $(y_{i1},y_{i2},\ldots,y_{i,Ti},u_i)$ is equal to the marginal distribution for u_i times the conditional distribution of $\mathbf{y}_i = (y_{i1},\ldots,y_{i,Ti})$ given u_i:

$$p(y_{i1},y_{i2},\ldots,y_{i,T_i},u_i\,|\,\mathbf{X}_i) = p(y_{i1},y_{i2},\ldots,y_{i,T_i}\,|\,\mathbf{X}_i,u_i)f(u_i),$$

where $f(u_i)$ is the marginal density for u_i. Now, we can obtain the marginal distribution of $(y_{i1},y_{i2},...,y_{i,Ti})$ without u_i by

$$p(y_{i1},y_{i2},...,y_{i,T_i}\,|\,\mathbf{X}_i)=\int_{u_i} p(y_{i1},y_{i2},...,y_{i,T_i}\,|\,\mathbf{X}_i,u_i)f(u_i)du_i.$$

For the specific application, with the Poisson conditional distributions for $y_{it}|u_i$ and a normal distribution for the random effect,

$$p(y_{i1},y_{i2},...,y_{i,T_i}\,|\,\mathbf{X}_i)=\int_{-\infty}^{\infty}\left[\prod_{t=1}^{T_i}\frac{\exp(-\mu_{it})\mu_{it}^{y_{it}}}{y_{it}!}\right]\frac{1}{\sigma}\phi\left(\frac{u_i}{\sigma}\right)du_i\,,\ \mu_{it}=\exp(\mathbf{x}_{it}'\boldsymbol{\beta}+u_i).$$

The log likelihood function will now be

$$\ln L=\sum_{i=1}^{n}\ln\left\{\int_{-\infty}^{\infty}\left[\prod_{t=1}^{T_i}\frac{\exp(-\mu_{it})\mu_{it}^{y_{it}}}{y_{it}!}\right]\frac{1}{\sigma}\phi\left(\frac{u_i}{\sigma}\right)du_i\right\},\ \mu_{it}=\exp(\mathbf{x}_{it}'\boldsymbol{\beta}+u_i).\quad(15\text{-}13)$$

The optimization problem is now free of the unobserved u_i, but that complication has been traded for another one, the integral that remains in the function.

To complete this part of the derivation, we will simplify the log likelihood function slightly in a way that will make it fit more naturally into the derivations to follow. Make the change of variable $u_i = \sigma w_i$ where w_i has mean zero and standard deviation one. Then, the Jacobian is $du_i = \sigma dw_i$, and the limits of integration for w_i are the same as for u_i. Making the substitution and multiplying by the Jacobian, the log likelihood function becomes

AU: Are 'd's" ok Roman?

$$\ln L=\sum_{i=1}^{n}\ln\left\{\int_{-\infty}^{\infty}\left[\prod_{t=1}^{T_i}\frac{\exp(-\mu_{it})\mu_{it}^{y_{it}}}{y_{it}!}\right]\phi(w_i)dw_i\right\},\ \mu_{it}=\exp(\mathbf{x}_{it}'\boldsymbol{\beta}+\sigma w_i).\quad(15\text{-}14)$$

The log likelihood is then maximized over $(\boldsymbol{\beta},\sigma)$. The purpose of the simplification is to parameterize the model so that the distribution of the variable that is being integrated out has no parameters of its own. Thus, in (15-14), w_i is normally distributed with mean zero and variance one.

In the next section, we will turn to how to compute the integrals. Section 14.9.6.c analyzes this model and suggests the **Gauss-Hermite quadrature** method for computing the integrals. In this section, we will derive a method based on simulation, **Monte Carlo integration**.[4]

FN 4

[4] The term "Monte Carlo" is in reference to the casino at Monte Carlo, where random number generation is a crucial element of the business.

15-23

15.6.2 Monte Carlo Integration

Note double quotes

Integrals often appear in econometric estimators in open form, that is, in a form for which there is no specific form function that is equivalent to them. (E.g., the integral, $\int_0^t \theta\exp(-\theta w)dw = 1-\exp(-\theta t)$, is in closed form. The integral in (15-14) is in open form.) There are various devices available for approximating open-form integrals. Gauss-Hermite and Gauss-Laguerre quadrature noted in Section 14.9.6.c and in Appendix E2.4 are two. The technique of Monte Carlo integration can often be used when the integral is in the form

$$h(y) = \int_w g(y\,|\,w)f(w)dw = E_w[g(y\,|\,w)]$$

where $f(w)$ is the density of w and and w is a random variable that can be simulated. [There are some necessary conditions on w and $g(y|w)$ that will be met in the applications that interest us here. Some details appear in Cameron and Trivedi (2005) and Train (2003).]

If $w_1, w_2, \ldots, w_n$ are a random sample of observations on the random variable w and $g(w)$ is a function of w with finite mean and variance, then by the law of large numbers [Theorem D.4 and the corollary in (D-5)],

$$\operatorname{plim}\frac{1}{n}\sum_{i=1}^{n} g(w_i) = E[g(w)].$$

The function in (15-14) is in this form;

$$\int_{-\infty}^{\infty}\left[\prod_{t=1}^{T_i}\frac{\exp[-\exp(\mathbf{x}_{it}'\boldsymbol{\beta}+\sigma w_i)][\exp(\mathbf{x}_{it}'\boldsymbol{\beta}+\sigma w_i)]^{y_{it}}}{y_{it}!}\right]\phi(w_i)dw_i$$
$$= E_{w_i}\left[g(y_{i1}, y_{i2}, \ldots, y_{iT_i}\,|\,w_i, \mathbf{X}_i, \boldsymbol{\beta}, \sigma)\right]$$

where

$$g(y_{i1}, y_{i2}, \ldots, y_{iT_i}\,|\,w_i, \mathbf{X}_i, \boldsymbol{\beta}, \sigma) = \prod_{t=1}^{T_i}\frac{\exp[-\exp(\mathbf{x}_{it}'\boldsymbol{\beta}+\sigma w_i)][\exp(\mathbf{x}_{it}'\boldsymbol{\beta}+\sigma w_i)]^{y_{it}}}{y_{it}!}$$

and w_i is a random variable with standard normal distribution. It follows, then, that

$$\operatorname{plim}\frac{1}{R}\sum_{r=1}^{R}\prod_{t=1}^{T_i}\frac{\exp[-\exp(\mathbf{x}_{it}'\boldsymbol{\beta}+\sigma w_{ir})][\exp(\mathbf{x}_{it}'\boldsymbol{\beta}+\sigma w_{ir})]^{y_{it}}}{y_{it}!}$$
$$= \int_{-\infty}^{\infty}\left[\prod_{t=1}^{T_i}\frac{\exp[-\exp(\mathbf{x}_{it}'\boldsymbol{\beta}+\sigma w_i)][\exp(\mathbf{x}_{it}'\boldsymbol{\beta}+\sigma w_i)]^{y_{it}}}{y_{it}!}\right]\phi(w_i)dw_i. \quad (15\text{-}15)$$

This suggests the strategy for computing the integral. We can use the methods developed in Section 15.2 to produce the necessary set of random draws on w_i from the standard normal distribution, then compute the approximation to the integral according to (15-15).

and

15-24

Example 15.8 Fractional Moments of the Truncated Normal Distribution

The following function appeared in Greene's (1990) study of the stochastic frontier model:

$$h(M,\varepsilon)=\frac{\int_0^{\infty} z^M \frac{1}{\sigma}\phi\left[\frac{z-(-\varepsilon-\theta\sigma^2)}{\sigma}\right]dz}{\int_0^{\infty} \frac{1}{\sigma}\phi\left[\frac{z-(-\varepsilon-\theta\sigma^2)}{\sigma}\right]dz}.$$

The integral only exists in closed form for integer values of M. However, the weighting function that appears in the integral is of the form

minus

$$f(z\,|\,z>0)=\frac{f(z)}{\text{Prob}[z>0]}=\frac{\frac{1}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right)}{\int_0^{\infty}\frac{1}{\sigma}\phi\left(\frac{z-\mu}{\sigma}\right)dz}.$$ minus

This is a truncated normal distribution. It is the distribution of a normally distributed variable z with mean μ and standard deviation σ, conditioned on z being greater than zero. The integral is equal to the expected value of z^M given that z is greater than zero when z is normally distributed with mean $\mu = -\varepsilon - \theta\sigma^2$ and variance σ^2. minus

The truncated normal distribution is examined in Section 18.2. The function $h(M,\varepsilon)$ is the expected value of z^M when z is the truncation of of a normal random variable with mean μ and standard deviation σ. To evaluate the integral by Monte Carlo integration, we would require a sample $z_1,\ldots,z_R$ from this distribution. We have the results we need in (15-4) with $L = 0$ so $P_L = \Phi[0-(-\varepsilon-\theta\sigma^2)/\sigma] = \Phi(\varepsilon/\sigma + \theta\sigma)$ and $U = +\infty$ so $P_U = 1$. Then, a draw on z is obtained by

minus minus

$$z = \mu + \sigma\Phi^{-1}[P_L + F(1-P_L)].$$

where F is the primitive draw from $U[0,1]$. Finally, the integral is approximated by the simple average of the draws,

$$h(M,\varepsilon) \approx \frac{1}{R}\sum_{r=1}^{R} z[\varepsilon,\theta,\sigma,F_r]^M.$$

This

~~The preceding~~ is an application of Monte Carlo integration. In certain cases, an integral can be approximated by computing the sample average of a set of function values. The approach taken here was to interpret the integral as an expected value. Our basic statistical result for the behavior of sample means implies that with a large enough sample, we can approximate the integral as closely as we like. The general approach is widely applicable in Bayesian econometrics and has begun to appear in classical statistics and econometrics as well.[5]

FN 5

[5] See Geweke (1986, 1988, 1989, 2005) for discussion and applications. A number of other references are given in Poirier (1995, p. 654) and Koop (2003).

15-25

weighting function. But now the range of integration is not $-\infty$ to $+\infty$; it is $-\mu/\sigma$ to $+\infty$. There is another approach. Suppose that z is a random variable with $N[\mu, \sigma^2]$ distribution. Then the density of the truncated normal (at zero) distribution for z is

$$f(z \mid z > 0) = \frac{f(z)}{\text{Prob}[z > 0]} = \frac{\frac{1}{\sigma}\phi\left[\frac{z-\mu}{\sigma}\right]}{\int_0^\infty \frac{1}{\sigma}\phi\left[\frac{z-\mu}{\sigma}\right] dz}.$$

This result is exactly the weighting function that appears in $h(r, \varepsilon)$, and the function being weighted is z^r. Therefore, $h(r, \varepsilon)$ is the expected value of z^r given that z is greater than zero. That is, $h(r, \varepsilon)$ is a possibly fractional moment—we do not restrict r to integer values—from the truncated (at zero) normal distribution when the untruncated variable has mean $-(\varepsilon + \theta\sigma^2)$ and variance σ^2.

Now that we have identified the function, how do we compute it? We have already concluded that the familiar quadrature methods will not suffice. (And, no one has previously derived closed forms for the fractional moments of the normal distribution, truncated or not.) But, if we can draw a random sample of observations from this truncated normal distribution $\{z_i\}$, then the sample mean of $w_i = z_i^r$ will converge in probability (mean square) to its population counterpart. [The remaining detail is to establish that this expectation is finite, which it is for the truncated normal distribution; see Amemiya (1973).] Because we showed earlier how to draw observations from a truncated normal distribution, this remaining step is simple.

The preceding is a fairly straightforward application of **Monte Carlo integration.** In certain cases, an integral can be approximated by computing the sample average of a set of function values. The approach taken here was to interpret the integral as an expected value. We then had to establish that the mean we were computing was finite. Our basic statistical result for the behavior of sample means implies that with a large enough sample, we can approximate the integral as closely as we like. The general approach is widely applicable in Bayesian econometrics and has begun to appear in classical statistics and econometrics as well.[3]

15.6.2.a ~~17.3.3~~ **HALTON SEQUENCES AND RANDOM DRAWS FOR SIMULATION-BASED INTEGRATION**

Monte Carlo integration is used to evaluate the expectation

$$E[g(x)] = \int_x g(x) f(x)\, dx$$

where $f(x)$ is the density of the random variable x and $g(x)$ is a smooth function. The Monte Carlo approximation is

$$\widehat{E[g(x)]} = \frac{1}{R}\sum_{r=1}^{R} g(x_r).$$

Convergence of the approximation to the expectation is based on the law of large numbers—a random sample of draws on $g(x)$ will converge in probability to its expectation. The standard approach to simulation-based integration is to use random draws from the specified distribution. Conventional simulation-based estimation uses

[3] See Geweke (1986, 1988, 1989) for discussion and applications. A number of other references are given in Poirier (1995, p. 654).

15-26

a random number generator to produce the draws from a specified distribution. The central component of this approach is draws from the standard continuous uniform distribution, $U[0, 1]$. Draws from other distributions are obtained from these draws by using transformations. In particular, for a draw from the normal distribution, where u_i is one draw from $U[0, 1]$, $v_i = \Phi^{-1}(u_i)$. Given that the initial draws satisfy the necessary assumptions, the central issue for purposes of specifying the simulation is the number of draws. Good performance in this connection requires very large numbers of draws. Results differ on the number needed in a given application, but the general finding is that when simulation is done in this fashion, the number is large (hundreds or thousands). A consequence of this is that for large-scale problems, the amount of computation time in simulation-based estimation can be extremely large. Numerous methods have been devised for reducing the numbers of draws needed to obtain a satisfactory approximation. One such method is to introduce some autocorrelation into the draws a small amount of negative correlation across the draws will reduce the variance of the simulation. **Antithetic draws**, whereby each draw in a sequence is included with its mirror image (w_i and $-w_i$ for normally distributed draws, w_i and $1-w_i$ for uniform, for example) is one such method. [See Geweke (1988) and Train (2009, Chapter 9).]

drawn/ minus KT

Procedures have been devised in the numerical analysis literature for taking "intelligent" draws from the uniform distribution, rather than random ones. [See Train (1999, 2009) and Bhat (1999) for extensive discussion and further references.] An emerging literature has documented dramatic speed gains with no degradation in simulation performance through the use of a smaller number of **Halton draws** or other constructed, nonrandom sequences instead of a large number of random draws. These procedures appear vastly to reduce the number of draws needed for estimation (sometimes by a factor of 90% or more) and reduce the simulation error associated with a given number of draws. In one application of the method to be discussed here, Bhat (1999) found that 100 Halton draws produced lower simulation error than 1,000 random numbers.

KT percent/

AV: Term "Halton draws" is not in chap. list

A sequence of Halton draws is generated as follows: Let r be a prime number. Expand the sequence of integers $g = 1, 2, \ldots$ in terms of the base r as

$$g = \sum_{i=0}^{I} b_i r^i \text{ where, by construction, } 0 \leq b_i \leq r - 1 \text{ and } r^I \leq g < r^{I+1}.$$

The Halton sequence of values that corresponds to this series is

$$H(g) = \sum_{i=0}^{I} b_i r^{-i-1}.$$

For example, using base 5, the integer 37 has $b_0 = 2$, $b_1 = 2$, and $b_3 = 1$. Then

$$H_5(37) = 2 \times 5^{-1} + 2 \times 5^{-2} + 1 \times 5^{-3} = 0.488.$$

The sequence of Halton values is efficiently spread over the unit interval. The sequence is not random as the sequence of pseudo-random numbers is; it is a well-defined deterministic sequence. But, randomness is not the key to obtaining accurate approximations to integrals. Uniform coverage of the support of the random variable is the central requirement. The large numbers of random draws are required to obtain smooth and dense coverage of the unit interval. Figures 15.3 and 15.4 show two sequences of 1,000 Halton draws and two sequences of 1,000 pseudo-random draws. The Halton draws are based on $r = 7$ and $r = 9$. The clumping evident in the first figure is the feature (among others) that mandates large samples for simulations.

Figs 15.3 15.4

15-27

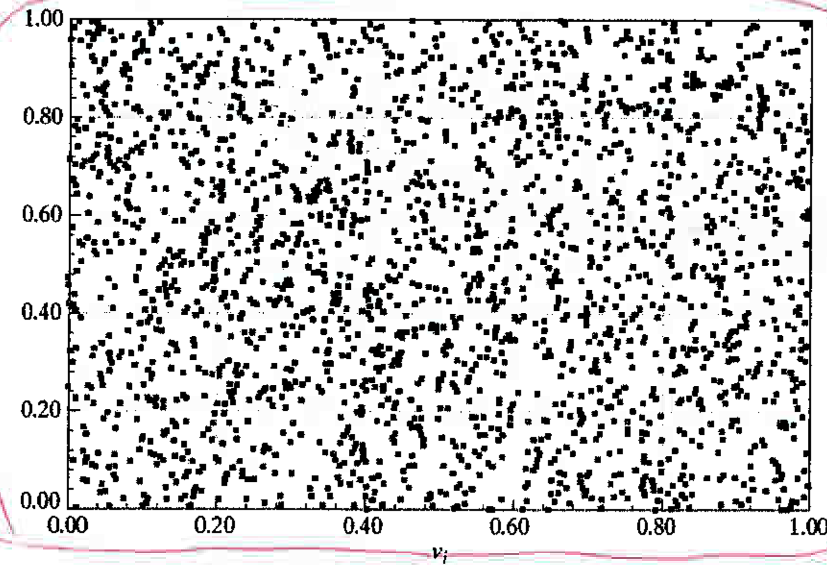

FIGURE ~~17.1~~ 15.3 Bivariate Distribution of Random Uniform Draws.

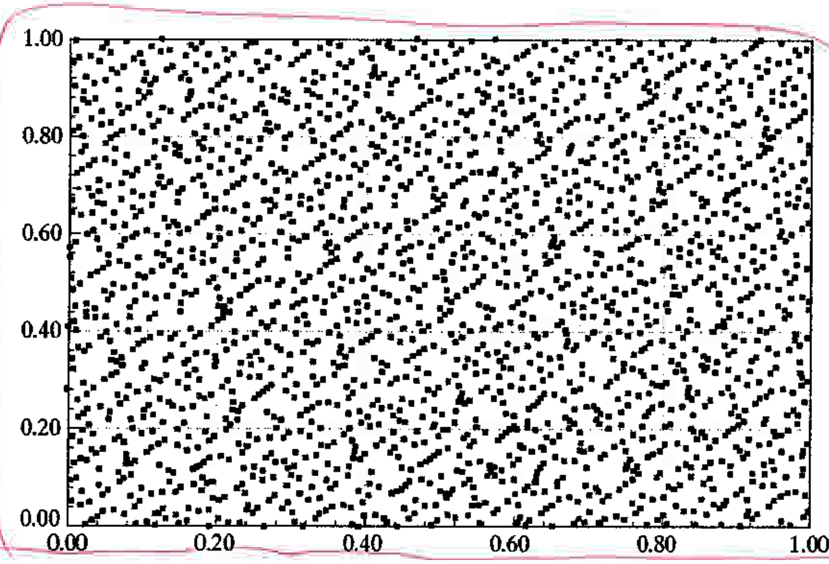

FIGURE ~~17.2~~ 15.4 Bivariate Distribution of Halton (7) and Halton (9).

Example ~~17.2~~ 15.9 Estimating the Lognormal Mean

We are interested in estimating the mean of a standard lognormally distributed variable. Formally, this is

$$E[y] = \int_{-\infty}^{\infty} \exp(x) \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}x^2\right] dx = 1.649.$$