Introduction to the Mathematical and Statistical Foundations of Econometrics

Remaining corrections and improvements in the 2004 and 2007 editions[1]

October 21, 2009

Page 8, section 1.2.3: Some of my students had difficulties understanding the derivation of the binomial probabilities in section 1.2.3. Here is a slightly different explanation:

Consider a bowl containing N balls of equal size and weight, of which K are red, and the remaining $N-K$ balls are white. Suppose you draw randomly n balls from this bowl *with replacement*, i.e., shake the bowl, pick a ball while blindfolded, take the blindfold off and write down the color of the ball, put the ball back in the bowl, and repeat this n times. The problem is: What is the probability that this sample contains exactly k red balls?

To answer this question, consider first the case that also the order in which the red balls are drawn matters. Thus, let us consider an event of the following type. The sample contains k red balls in positions $i_1 < i_2 < \cdots < i_k \le n$. For example, the case $n = 3$, $k = 2$, $i_1 = 1$, $i_2 = 3$ corresponds to (*red, white, red*).

In general, there are K ways to put a red ball in each of the positions $i_1 < i_2 < \cdots < i_k$, so that there are K^k different ways to fill the positions $i_1 < i_2 < \cdots < i_k$ with red balls. Similarly, there are $(N-K)^{n-k}$ ways to fill the remaining $n-k$ positions with white balls. Thus, the total number of ways we can draw, with replacement, a sample of size n containing k red balls in positions $i_1 < i_2 < \cdots < i_k$ is $K^k(N - K)^{n-k}$. For example, in the case $n = 3$, $k = 2$ the number of ways we can draw each of the samples (*red, white, red*), (*white, red, red*), (*red, red, white*) with replacement is $K^k(N - K)^{n-k}$. Moreover, similar to the Texas lotto case it follows that the number of *unordered* sets of k red balls and n-k white balls is: n choose k, which in the case n = 3, $k = 2$ yields $\binom{3}{2} = 3$. Thus, the total number of ways we can choose a sample of size n *with replacement* containing k red balls in any order is:

$$\binom{n}{k}K^k(N - K)^{n-k}.$$

Finally, the number of ways we can choose a sample of size n with replacement is N^n.

---

[1] Page and line numbers refer to the 2007 edition. Thanks to Dongkoo Kim and Soo Hyun Oh for pointing out many errors.

Therefore, the probability that we draw a sample *with replacement* containing k read balls and n-k white balls in *any order* from a bowl containing K red balls and $N-K$ white balls is:

$$P(\{k\}) = \binom{n}{k}\frac{K^k(N-K)^{n-k}}{N^n} = \binom{n}{k}p^k(1-p)^{n-k}, \quad k = 0,1,2,....,n, \tag{1.15}$$

$$\text{where } p = K/N.$$

The sample space Ω and the σ-algebra $\mathscr{F}$ in this case are the same as in the sampling without replacement case, i.e., $\Omega = \{0,1,2,....,n\}$ and $\mathscr{F}$ is the collection of all subsets of Ω, including the empty set $\emptyset$ and Ω itself. The only difference is the probability measure P. In particular, any set A in $\mathscr{F}$ is either empty or takes the form $A = \{k_1, k_2, ..., k_m\}$, where $0 \le k_1 < k_2 < \cdots < k_m \le n$ and $m \le n$. In the latter case $P(A) = \sum_{j=1}^m P(\{k_j\})$, where $P(\{k\})$ is now defined by (1.15).

Page 9, line 2 from top: Replace the first line of the equation with

$$P(\{k\}) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} = \frac{\dfrac{K!(N-K)!}{\mathbf{k!}(K-k)!(n-k)!(N-K-n+k)!}}{\dfrac{N!}{n!(N-n)!}}$$

(the actual correction is the bold **k!**)

Page 10, line 3 from below: Replace $[0.5-1/q, 0.5+1/q]$ with $(0.5-1/q, 0.5+1/q]$

Page 11, line 2 from top: Replace $[0.5-1/q, 0.5+1/q]$ with $(0.5-1/q, 0.5+1/q]$

Page 11, line 4 from top: Replace $[0.5-1/q, 0.5+1/q]$ with $(0.5-1/q, 0.5+1/q]$

Page 18, line 8 from top: Replace "probability" with "unique probability"

Page 18, line 10 from top: Delete this line

Page 21, line 5 of Definition 1.8: Replace $\mathscr{B}^{k}$. with $\mathscr{B}^{k}$,

Page 25, line 10 from top: Replace $\mu([-\infty, a))$ with $\mu((-\infty, a))$

Page 31, exercise 16: Replace "Definition 1.12" with "Definition 1.11"

Page 36, lines 6 and 7: Replace these two lines with the following:[2]

Proof: Suppose that there exist two extensions of outer measures, P_1 and P_2, to probability measures on $\{\Omega, \mathscr{F}\}$ such that for all sets $B \in \mathscr{F}_0$, $P_1(B) = P_2(B) = P(B)$. Then by the definition of outer measure, for all sets $A \in \mathscr{F}$,

$$\begin{aligned} P_1(A) &= \inf_{A \subset \cup_{i=1}^{\infty} B_i,\ B_i \in \mathscr{F}_0} \textstyle\sum_{i=1}^{\infty} P(B_i) = \inf_{A \subset \cup_{i=1}^{\infty} B_i,\ B_i \in \mathscr{F}_0} \textstyle\sum_{i=1}^{\infty} P_2(B_i) \\ &\geq \inf_{A \subset \cup_{i=1}^{\infty} B_i,\ B_i \in \mathscr{F}_0} P_2\left(\cup_{i=1}^{\infty} B_i\right) \geq \inf_{A \subset \cup_{i=1}^{\infty} B_i,\ B_i \in \mathscr{F}_0} P_2(A) \\ &= P_2(A) \end{aligned}$$

and similarly, $P_2(A) \geq P_1(A)$. Thus, $P_2(A) = P_1(A)$ for all sets $A \in \mathscr{F}$. Q.E.D.

Page 41, line 16 from top: Replace $\left(\inf_{x \in B(j,m,n) \atop *} g(x_*)\right)$ with $\left(\inf_{x_* \in B(j,m,n)} g(x_*)\right)$

Page 42, line 1 of section 2.3: Replace $\sum_{j=1}^{m}$ with $\sum_{j=0}^{m}$

Page 42, line 3 of section 2.3: Replace $\sum_{j=1}^{m}$ with $\sum_{j=0}^{m}$ twice

Page 43, line 1 of Definition 2.5: Replace $\{\mathbb{R}^k \mathscr{B}^k\}$ with $\{\mathbb{R}^k, \mathscr{B}^k\}$

Page 45, equation (2.11): Replace $\leq$ with $<$

Page 46, line 2 from below: Replace $X(\Omega)$ with $X(\omega)$

Page 47, line 2 of Definition 2.9: Replace "of *P*" with "to *P*"

Page 49: Replace the proof of Theorem 2.18 with the following:

Let $g(x)$ be a nonnegative Borel measurable function on $\mathbb{R}$, and let X be a random variable defined on the probability space $\{\Omega, \mathscr{F}, P\}$, with induced probability measure $\mu_X(.)$ defined on the Borel sets in $\mathbb{R}$. Let $S_n(g)$ be the set of all nonnegative simple functions

---

[2] It was stated that the proof of Lemma 1.B.4. is too difficult and too long. However, there is a much shorter proof in

Jeffrey S. Rosenthal: *A First Look at Rigorous Probability Theory* (2000), World Scientific Publishing Company, p.14.

Thanks to David Jinkins for pointing this out to me.

$g_n(x) = \sum_{j=1}^n a_j I(x \in B_j)$ with $a_j \le \inf_{x \in B_j} g(x)$, so that $g_n(x) \le g(x)$ pointwise in x. Without loss of generality we may assume that $0 = a_1 < a_2 < ..... < a_n$ and $\cup_{j=1}^n B_j = \mathbb{R}$. Then by definition,

$$\int g(x)d\mu_X(x) = \sup_{g_* \in \cup_{n=1}^\infty S_n(g)} \int g_*(x)d\mu_X(x).$$

Let $Y = g(X)$, and denote by $S_n(Y)$ be the set of all nonnegative simple random variables $Y_n(\omega) = \sum_{j=1}^n b_j I(\omega \in A_j)$ with $b_j \le \inf_{\omega \in A_j} Y(\omega)$, so that $Y_n(\omega) \le Y(\omega)$ pointwise in $\omega \in \Omega$. Again, without loss of generality we may assume that $0 = b_1 < b_2 < ..... < b_n$ and $\cup_{j=1}^n A_j = \Omega$. Then by definition,

$$\int g(X(\omega))dP(\omega) = \int Y(\omega)dP(\omega) = \sup_{Y_* \in \cup_{n=1}^\infty S_n(Y)} \int Y_*(\omega)dP(\omega),$$

Each simple function $g_n(x) = \sum_{j=1}^n a_j I(x \in B_j) \in S_n(g)$ corresponds to a simple random variable

$$Y_n^*(\omega) = g_n(X(\omega)) = \sum_{j=1}^n a_j I(X(\omega) \in B_j) = \sum_{j=1}^n a_j I(\omega \in C_j) \in S_n(Y)$$

where $C_j = \{\omega \in \Omega: X(\omega) \in B_j\} \in \mathscr{F}$, such that

$$\int Y_n^*(\omega)dP(\omega) = \sum_{j=1}^n a_j P(C_j) = \sum_{j=1}^n a_j \mu_X(B_j) = \int g_n(x)d\mu_X(x).$$

Hence,

$$\int g_n(x)d\mu_X(x) \le \sup_{Y_* \in \cup_{n=1}^\infty S_n(Y)} \int Y_*(\omega)dP(\omega) = \int Y(\omega)dP(\omega)$$

and thus,

$$\int g(x)d\mu_X(x) = \sup_{g_* \in \cup_{n=1}^\infty S_n(g)} \int g_*(x)d\mu_X(x) \le \int Y(\omega)dP(\omega).$$

On the other hand, each simple random variable $Y_n(\omega) = \sum_{j=1}^n b_j I(\omega \in A_j) \in S_n(Y)$ corresponds to a simple function $g_n^*(x) = \sum_{j=1}^n b_j I(x \in B_j) \in S_n(g)$ such that

$$g_n^*(X(\omega)) = \sum_{j=1}^n b_j I(X(\omega) \in B_j) \ge Y_n(\omega)$$

To see this, let

$$B_j = \{x \in \mathbb{R}: g(x) \in [b_j, b_{j+1})\} \text{ for } j = 1,...,n-1,\ B_n = \{x \in \mathbb{R}: g(x) \in [b_n, \infty)\}.$$

Then

$$g_n^*(X(\omega)) = \sum_{j=1}^{n} b_j I(X(\omega) \in B_j) = \sum_{j=1}^{n-1} b_j I(b_j \le g(X(\omega)) < b_{j+1}) + b_n I(g(X(\omega)) \ge b_n)$$

$$= \sum_{j=1}^{n-1} b_j I(b_j \le Y(\omega) < b_{j+1}) + b_n I(Y(\omega) \ge b_n)$$

$$\ge \sum_{j=1}^{n-1} b_j I(b_j \le Y_n(\omega) < b_{j+1}) + b_n I(Y_n(\omega) \ge b_n)$$

$$= \sum_{j=1}^{n} b_j I(Y_n(\omega) = b_j) = \sum_{j=1}^{n} b_j I(\omega \in A_j) = Y_n(\omega).$$

Hence,

$$\int Y_n(\omega)dP(\omega) \le \int g_n^*(X(\omega))dP(\omega) = \int g_n^*(x)d\mu_X(x)$$

$$\le \sup_{g_* \in \cup_{n=1}^{\infty} S_n(g)} \int g_*(x)d\mu_X(x) = \int g(x)d\mu_X(x)$$

and thus

$$\int Y(\omega)dP(\omega) = \sup_{Y* \in \cup_{n=1}^{\infty} S_n(Y)} \int Y_n(\omega)dP(\omega) \le \int g(x)d\mu_X(x).$$

Consequently,

$$\int g(X(\omega))dP(\omega) = \int Y(\omega)dP(\omega) = \int g(x)d\mu_X(x) \qquad (2.16)$$

for all nonnegative Borel measurable functions $g(x)$, and therefore (2.16) holds also for all Borel measurable functions $g(x)$. Q.E.D.

Page 50, line 1 of Definition 2:13: Replace *m's* with *m-th*

Page 57, line 2 from top: Replace $E[\varphi(X)]$ with $E[\varphi(Y)]$

Page 57, line 15 from top: Replace P_j with p_j

Page 59, line 1 of exercise 17: Replace $E(E)$ with $E(X)$

Page 74, Theorem 3.5: Replace [3] $P[E(Y|\mathscr{F}) = E(Y)] = 1$ with $E[Y|\mathscr{F}] \equiv E[Y]$

Page 74, Proof of Theorem 3.5: Replace the proof with the following:

Denote $Z = E[Y|\mathscr{F}]$. It is left as an exercise to prove that $Z = E[Y]$ a.s., along the same lines as the proofs of Theorems 3.2 and 3.4. Next, recall from Definition 3.1 that Z is measurable $\mathscr{F}$, i.e., for any Borel set B the set $\{\omega \in \Omega\colon Z(\omega) \in B\}$ is contained in $\mathscr{F}$. Hence, for any Borel set B we have either $\{\omega \in \Omega\colon Z(\omega) \in B\} = \Omega$ or $\{\omega \in \Omega\colon Z(\omega) \in B\} = \emptyset$. Now let B be the singleton $\{E[Y]\}$. Then either $\{\omega \in \Omega\colon Z(\omega) = E[Y]\} = \Omega$ or $\{\omega \in \Omega\colon Z(\omega) = E[Y]\} = \emptyset$.

[3] Thanks to Renxiang Dai, Tilburg University, for suggesting this.

The latter is excluded by $Z = E[Y]$ a.s. , hence $Z(\omega) \equiv E[Y]$ for all $\omega \in \Omega$. In other words, $Z = E[Y]$ holds exactly.

Page 83, Exercise 10: Replace "Borel-measurable" with "continuous"

Page 89, line 9 from below: Replace the expression for $\varphi_{NB(m,p)}(t)$ with

$$\begin{aligned}\varphi_{NB(m,p)}(t) &= \left(\frac{p}{1 - (1 - p)e^{i.t}}\right)^m \\ &= \left(\frac{p}{1 - (1 - p)\cos(t) - i.(1 - p)\sin(t)}\right)^m \\ &= \left(\frac{p\left(1 - (1 - p)\cos(t) + i.(1 - p)\sin(t)\right)}{p^2 + 2(1 - p)(1 - \cos(t))}\right)^m\end{aligned}$$

Page 101, line 9 from below: Replace the expression for $\varphi_{U[a,b]}(t)$ with

$$\begin{aligned}\varphi_{U[a,b]}(t) &= \frac{\exp(i.b.t) - \exp(i.a.t)}{i.(b - a)t} \\ &= \frac{(\sin(b.t) - \sin(a.t)) - i(\cos(b.t) - \cos(a.t))}{(b - a)t}\end{aligned}$$

Page 115, line 3 from below: Replace $\begin{pmatrix} 1 & B^T \\ 0 & I_k \end{pmatrix}$ with $\begin{pmatrix} 1 & \beta^T \\ 0 & I_k \end{pmatrix}$

Page 141, equation (6.2): Replace $\sum_{k=1}^{j-1}$ with $\sum_{k=1}^{j}$ twice

Page 141, line 2 below eq. (6.2): Replace $\Sigma_{k=1}^{j-1}$ with $\Sigma_{k=1}^{j}$

Page 143, Proof of Theorem 6.5: Replace the sentence “Again, without loss of generality we may assume that P[X = 0] = 1 and that X_n is a nonnegative random variable.” with the following:

It will be shown first that $X_n \rightarrow_p X$ and $\lim_{M\to\infty}\sup_{n\geq 1}E[|X_n|I(|X_n| > M)] = 0$ imply that $|X_n - X|$ is uniformly integrable. To show this, we need to show that $\sup_{n\geq 1}E[|X_n|] < \infty$ and $E[|X|] < \infty$, as follows. Let $M > 0$. Then

$$E[|X_n|] = E[|X_n|.I(|X_n| \leq M)] + E[|X_n|.I(|X_n| > M)] \leq M + \sup_{n\geq 1}E[|X_n|.I(|X_n| > M)$$

Because X_n is uniformly integrable, for an arbitrary $\varepsilon > 0$ we can choose an M_ε such that the second term is less than ε. Consequently, $\sup_{n\geq 1}E[|X_n|] \leq M_\varepsilon + \varepsilon < \infty$.

To show that $E[|X|] < \infty$, choose $K > 0$ and $\varepsilon > 0$ arbitrary. Then

$$E[|X|.I(|X| \leq K)] = E[|X|.I(|X_n - X| > \varepsilon).I(|X| \leq K)] + E[|X|.I(|X_n - X| \leq \varepsilon).I(|X| \leq K)]$$

$$\leq K.P[|X_n - X| > \varepsilon] + E[|X_n - X|.I(|X_n - X| \leq \varepsilon).I(|X| \leq K)]$$

$$+ E[|X_n|.I(|X_n - X| \leq \varepsilon).I(|X| \leq K)]$$

$$\leq K.P[|X_n - X| > \varepsilon] + \varepsilon + \sup_{m\geq 1}E[|X_m|]$$

Letting $n \to \infty$ it follows from $X_n \to_p X$ that $E[|X|.I(|X| \leq K)] \leq \varepsilon + \sup_{m\geq 1}E[|X_m|]$ and next, letting $K \to \infty$, it follows that $E[|X|] \leq \varepsilon + \sup_{n\geq 1}E[|X_n|]$. Because ε was arbitrary, it follows now that $E[|X|] \leq \sup_{n\geq 1}E[|X_n|] < \infty$.

To show that $|X_n - X|$ is uniformly integrable, observe that for arbitrary $M > 0$ and $K > 0$,

$$E[|X_n - X|.I(|X_n - X| > M)] = E[|X_n - X|.I(|X_n - X| > M).I(|X_n| \leq K).I(|X| \leq K)]$$

$$+ E[|X_n - X|.I(|X_n - X| > M).I(|X_n| > K).I(|X| \leq K)]$$

$$+ E[|X_n - X|.I(|X_n - X| > M).I(|X_n| \leq K).I(|X| > K)]$$

$$+ E[|X_n - X|.I(|X_n - X| > M).I(|X_n| > K).I(|X| > K)]$$

$$\leq 4.K.P[|X_n - X| > M] + 2\sup_{m\geq 1}E[|X_m|.I(|X_m| > K)]$$

$$+ 2.E[|X|.I(|X| > K)]$$

Hence,

$$\limsup_{n\to\infty}E[|X_n - X|.I(|X_n - X| > M)] \leq 2.\sup_{m\geq 1}E[|X_m|.I(|X_m| > K)] + 2.E[|X|.I(|X| > K)]$$

and thus, letting $K \to \infty$, it follows that $\lim_{n\to\infty}E[|X_n - X|.I(|X_n - X| > M)] = 0$.

Next, choose an arbitrary $\varepsilon > 0$, and pick an $M_0 > 0$. Then there exists a natural number $n_0(\varepsilon)$ such that $E[|X_n - X|.I(|X_n - X| > M_0)] < \varepsilon$ for all $n > n_0(\varepsilon)$, and therefore also

$$E[|X_n - X|.I(|X_n - X| > M)] < \varepsilon \text{ for all } n > n_0(\varepsilon) \text{ and } M > M_0.$$

Hence,

$$\lim_{M\to\infty}\sup_{n\geq 1}E[|X_n - X|.I(|X_n - X| > M)] \leq \varepsilon + \max_{1\leq n\leq n_0(\varepsilon)} \lim_{M\to\infty}E[|X_n - X|.I(|X_n - X| > M)]$$

$$= \varepsilon$$

where the equality follows from the fact that $E[|X_n - X|] < \infty$. Because ε was arbitrary, it follows now that $|X_n - X|$ is uniformly integrable. Therefore, without loss of generality we may now assume that $P[X = 0] = 1$ and that X_n is a nonnegative random variable.

Page 146, first line below (6.8): Replace Theorem 6.3 with Theorem 6.10

Page 147, equation (6.11): Replace equation (6.11) with:

$$E[g(X_1,\theta)] = \int_{-\infty}^{\infty} \frac{\exp(-(x+\theta_0-\theta)^2/2)/\sqrt{2\pi}}{\pi(1+x^2)}dx = \int_{-\infty}^{\infty} f(x-\theta+\theta_0)h(x|0)dx = \gamma(\theta-\theta_0), \tag{6.11}$$

Page 149, line 1 from top: Replace the sentence "*where B is a closed and bounded subset of* $\mathbb{R}^k$ *containing c*" with "*where B is an open subset of* $\mathbb{R}^k$ *containing c*" [4]

Page 159, line 7 from below: Replace Theorem 6.1 with Theorem 6.28

Page 161, last line: Replace Theorem 6.21 with Theorem 6.12

Page 185, line 3 of Theorem 7.5: Replace $\varnothing$ with 0

Page 217: Replace Assumption 8.2 with:

Assumption 8.2: $\text{plim}_{n\to\infty}\sup_{\theta\in\Theta}|n^{-1}\ln(\hat{L}_n(\theta)) - E[n^{-1}\ln(\hat{L}_n(\theta))]| = 0$ *and* $\lim_{n\to\infty}\sup_{\theta\in\Theta}|n^{-1}E[\ln(\hat{L}_n(\theta)/\hat{L}_n(\theta_0))] - \ell(\theta|\theta_0)| = 0$, *where* $\ell(\theta|\theta_0)$ *is a continuous function of* θ *such that for arbitrarily small* $\delta > 0$, $\sup_{\theta\in\Theta:\ \|\theta-\theta_0\|\geq\delta}\ell(\theta|\theta_0) < 0$.

Page 244, line 16 from top: Replace $P_{i,j}P_{j,i} = 1$ with $P_{i,j}P_{j,i} = I$

Page 253, line 1 of Section I.8: Replace "Theorem I.9" with "Theorem I.11"

---

[4] The reason is that the singleton $B = \{c\}$ is closed and bounded, and that for this case Theorem 6.12 may not hold. If B is open then there exists a $\delta > 0$ such that $\{x \in \mathbb{R}^k\colon \|x-c\| < \delta\} \subset B$, which is an essential element of the proof of Theorem 6.12. Alternatively, one may replace the sentence involved with "*where B is subset of* $\mathbb{R}^k$ *containing c in its interior*", as then B contains an open neighborhood of c.

Page 254, line 3 of Theorem I.12: Replace "Theorem I.9" with "Theorem I.11"

Page 254, line 3 of Theorem I.13: Replace "Theorem I.9" with "Theorem I.11"

Page 256, last line of Section I.8: Replace "Theorem I.5" with "Theorem I.15 "

Page 286, line 19 from top: Replace $A \subset B$ with $A \supset C$