Asymptotic coupon collector statistics when coupons are words

David Moews

January 18, 2006

Abstract. Given a probability distribution on a finite set of letters, we can construct a probability distribution on words of length L by assuming letters are independent. For large L, we find the asymptotic expected coupon-collector time for this distribution, and the limiting behavior of the trajectories of the number of coupons collected and the total probability mass covered.

We suppose that we are given a finite alphabet and a probability distribution on this alphabet; let the alphabet $\mathcal{A}$ be of size M, $\mathcal{A} = \{1, \ldots, M\}$, say, and let its letters have probabilities $q_1, \ldots, q_M$. To avoid trivial cases, assume that $M > 1$ and that each q_i is positive. Let the set of words of length L over this alphabet be W_L; W_L clearly has size $N_L = M^L$. We can obtain a probability distribution D_L on W_L by assuming that distinct letters are uncorrelated, i.e., that the probability of any given word with a_1 1s, ..., a_M Ms is $\prod_{1 \le i \le M} q_i^{a_i}$. Let $\{X_{L,t} \mid L \ge 0, t > 0\}$ be independent random variables, $X_{L,t}$ having distribution D_L. We are interested in C_L, the smallest t such that $\{X_{L,1}, \ldots, X_{L,t}\}$ is equal to W_L, and $N_{L,t}$, the number of distinct elements in $\{X_{L,1}, \ldots, X_{L,t}\}$; we are also interested in $P_{L,t}$, which we define to be the total probability mass covered by $X_{L,1}$, ..., $X_{L,t}$, i.e.,

$$P_{L,t} = \mathbf{P}(\text{there exists } i \in \{1, \ldots, t\} \text{ such that } X_{L,i} = Q \mid X_{L,1}, \ldots, X_{L,t}),$$

where Q has distribution D_L and is independent of the $X_{L,j}$'s.

Theorem 1 *Let μ be the geometric mean of the q_i's, and $t_1, t_2, \ldots \in \mathbf{Z}_{>0}$ be a sequence of times.*

1. *If the q_i's are all equal, write $t_L = \mu^{-L}\gamma_L$. Now if $\gamma_L \to \gamma$ as $L \to \infty$, then $P_{L,t_L} = N_{L,t_L}/N_L \to 1 - e^{-\gamma}$ as $L \to \infty$, in probability. (In fact, we have $\mathbf{E}N_{L,t_L}/N_L \to 1 - e^{-\gamma}$ and $\operatorname{Var} N_{L,t_L} \le \min(\mathbf{E}N_{L,t_L}, N_L - \mathbf{E}N_{L,t_L}) \le N_L$.)*

2. *If the q_i's are not all equal, let σ be the square root of $M^{-1} \sum_{1 \le i \le M} (\log q_i - \log \mu)^2$, and write $t_L = \mu^{-L} e^{\sqrt{L}\sigma\gamma_L}$. Now if $\gamma_L \to \gamma$ as $L \to \infty$, then*

$N_{L,t_L}/N_L \to \Phi(\gamma)$ as $L \to \infty$, in probability, where Φ is the distribution function of the standard normal distribution. (In fact, we have $\mathbf{E}N_{L,t_L}/N_L \to \Phi(\gamma)$ *and* $\mathrm{Var}\ N_{L,t_L} \leq \min(\mathbf{E}N_{L,t_L}, N_L - \mathbf{E}N_{L,t_L}) \leq N_L$*.)*

3. *If the q_i's are not all equal, let $\bar{\mu} = \prod_{1\leq i\leq M} q_i^{q_i}$, let $\bar{\sigma}$ be the square root of $\sum_{1\leq i\leq M} q_i(\log q_i - \log\bar{\mu})^2$, and write $t_L = \bar{\mu}^{-L}e^{\sqrt{L}\bar{\sigma}\bar{\gamma}_L}$. If $\bar{\gamma}_L \to \bar{\gamma}$ as $L \to \infty$, then $P_{L,t_L} \to \Phi(\bar{\gamma})$ as $L \to \infty$, in probability, where Φ is the distribution function of the standard normal distribution. (In fact, we have* $\mathbf{E}P_{L,t_L} \to \Phi(\bar{\gamma})$ *and* $\mathrm{Var}\ P_{L,t_L} \leq \max(q_1, \ldots, q_M)^L$*.)*

Proof. Consider a coupon-collector problem with probabilities $p_1, \ldots, p_n$, and let Z_t be the number of coupons collected at time t. Clearly $Z_t = Y_1 + \cdots + Y_n$, where Y_i is the indicator function of having collected coupon i at time t. Therefore $\mathbf{E}Z_t = \sum_{1\leq i\leq n}\mathbf{E}Y_i = \sum_{1\leq i\leq n} 1-(1-p_i)^t$. Also, if $i \neq j$, Y_i and Y_j are anticorrelated: $\mathbf{E}(Y_i \mid Y_j = 0) = 1-(1-p_i/(1-p_j))^t \geq \mathbf{E}Y_i$, so $\mathbf{E}(Y_i \mid Y_j = 1) \leq \mathbf{E}Y_i$ and $\mathbf{E}(Y_iY_j) \leq \mathbf{E}Y_i\mathbf{E}Y_j$. Therefore

$$\begin{aligned}
\mathrm{Var}\ Z_t &= \mathbf{E}(Z_t^2) - (\mathbf{E}Z_t)^2 \\
&\leq \sum_{1\leq i\leq n} \mathbf{E}(Y_i^2) - (\mathbf{E}Y_i)^2 \\
&= \sum_{1\leq i\leq n} (\mathbf{E}Y_i)(1-\mathbf{E}Y_i) \\
&= \sum_{1\leq i\leq n} (1-p_i)^t(1-(1-p_i)^t) \\
&\leq \min(\sum_{1\leq i\leq n} 1-(1-p_i)^t, \sum_{1\leq i\leq n}(1-p_i)^t) \\
&= \min(\mathbf{E}Z_t, n - \mathbf{E}Z_t).
\end{aligned}$$

Now, if we let W_t be the total probability mass covered by time t, then $W_t = p_1Y_1 + \cdots + p_nY_n$, so

$$\begin{aligned}
\mathrm{Var}\ W_t &= \mathbf{E}(W_t^2) - (\mathbf{E}W_t)^2 \\
&\leq \sum_{1\leq i\leq n} p_i^2(\mathbf{E}Y_i)(1-\mathbf{E}Y_i) \\
&\leq \max(p_1, \ldots, p_n).
\end{aligned}$$

This proves the claims about the variance. To prove part 1, observe that if the q_i's are all equal, then they must all equal $1/M$; therefore $\mu = 1/M$. Also, D_L will be the uniform distribution on $N_L = M^L$ elements, so $P_{L,t} = N_{L,t}/N_L$. In this case therefore the coupon-collector problem we are considering has $p_i = M^{-L}$ for all i, so

$$\mathbf{E}N_{L,t_L}/N_L = M^L(1-(1-M^{-L})^{t_L})/M^L = 1-(1-M^{-L})^{M^L\gamma_L},$$

which clearly converges to $1-e^{-\gamma}$. Convergence in probability now follows from the fact that the variance of $N_{L,t_L}/N_L$ converges to 0.

To prove part 2, return to the general coupon-collector problem. Let ν be the (uniform) measure on the set $\{\log p_1, \ldots, \log p_n\}$ which assigns measure $1/n$ to each $\log p_i$, and fix some $K > 0$. Now if $\log t > K - \log p_i$,

$$\mathbf{E}Y_i = 1 - (1 - p_i)^t > 1 - (1 - p_i)^{e^K/p_i} \geq 1 - e^{-e^K},$$

so

$$\mathbf{E}Z_t/n \geq (1 - e^{-e^K})\nu((K - \log t, \infty)). \tag{1}$$

Similarly, if $\log t \leq -1 - K - \log p_i$, then since $1 - \beta/e \geq e^{-\beta}$ for $0 \leq \beta \leq 1$,

$$\mathbf{E}Y_i = 1 - (1 - p_i)^t \leq 1 - (1 - e^{-(K+1)}/t)^t \leq 1 - e^{-e^{-K}},$$

so

$$\mathbf{E}Z_t/n \leq 1 - e^{-e^{-K}} + \nu((-1 - K - \log t, \infty)). \tag{2}$$

Now, in our coupon-collector problem for words of length L, $\nu = \nu_L$, say, is obtained by taking the uniform measure on $\{\log q_1, \ldots, \log q_M\}$ and raising it to the Lth convolution power. Therefore, by the Central Limit Theorem, if u_1, u_2, …is a sequence such that $(u_L - L\log\mu)/(\sigma\sqrt{L}) \to u$, then $\nu_L((u_L, \infty)) \to 1 - \Phi(u)$. If we substitute this estimate for ν into (1) and let K go to infinity slowly (for example, we could set $K = L^{1/4}$), we find that $\liminf_{L\to\infty} \mathbf{E}N_{L,t_L}/N_L \geq 1 - \Phi(-\gamma)$. Similarly, substituting into (2), $\limsup_{L\to\infty} \mathbf{E}N_{L,t_L}/N_L \leq 1 - \Phi(-\gamma)$. By symmetry, however, $1 - \Phi(-\gamma) = \Phi(\gamma)$, so this proves that $\mathbf{E}N_{L,t_L}/N_L \to \Phi(\gamma)$. Convergence in probability now follows as before.

For part 3, consider the general coupon-collector problem, let $\bar{\nu}$ be the measure on $\{\log p_1, \ldots, \log p_n\}$ which assigns measure p_i to each $\log p_i$, and fix $K > 0$. The same argument that gave (1) and (2) then gives

$$\mathbf{E}W_t \geq (1 - e^{-e^K})\bar{\nu}((K - \log t, \infty)) \tag{3}$$

and

$$\mathbf{E}W_t \leq 1 - e^{-e^{-K}} + \bar{\nu}((-1 - K - \log t, \infty)). \tag{4}$$

Specializing to words of length L, we find that $\bar{\nu}$ is obtained by starting with the measure on $\{\log q_1, \ldots, \log q_M\}$ which assigns probability q_i to $\log q_i$ and raising it to the Lth convolution power. The proof now proceeds as in part 2. ∎

Theorem 2 *Let $\underline{q}$ be the minimum of the q_i's, and assume that $\underline{q} = q_1$ and that r of the q_i's are equal to $\underline{q}$. Then as $L \to \infty$:*

1. *If $r > 1$, then $\mathbf{E}C_L = \underline{q}^{-L}(L\log r + O(1))$, and $\sqrt{\mathrm{Var}\ C_L}/\mathbf{E}C_L = \Theta(L^{-1})$.*

2. *If $r = 1$, let β be the supremum of $(b_2 + \cdots + b_M)\prod_{2\leq j\leq M}(\underline{q}/q_j)^{b_j}$ over all nonnegative integers b_2, …, b_M. Then $0 < \beta < \infty$, β is attained by some b_2, …, b_M, $\mathbf{E}C_L = \beta\underline{q}^{-L}(\log L + O(1))$, and $\sqrt{\mathrm{Var}\ C_L}/\mathbf{E}C_L = \Theta((\log L)^{-1})$.*

Comment. Let us divide words into types, leaving the division arbitrary for the moment, except that words of the same type must have the same probability. If a type has n words of probability p, the expected time to collect all words of the type will be $p^{-1}(1 + 1/2 + \cdots + 1/n)$, or approximately $p^{-1} \log n$. This is clearly a lower bound on $\mathbf{E}C_L$. During the proof, we will see that $\mathbf{E}C_L$ is in fact governed by the waiting time to collect all words of a given type, one for which $p^{-1} \log n$ is large. If $r > 1$, this type will be the type of words made up completely of letters whose probability is smallest; if $r = 1$, this type will be the type of words made up of $L - b_2 - \cdots - b_M$ 1s, b_2 2s, …, b_M Ms, for the maximizing $b_2, \ldots, b_M$ above.

Proof. Consider the coupon collector problem where we have c_1 coupons of type 1 with probability $p_1, \ldots, c_n$ coupons of type n with probability p_n, and let D be the time until we collect all coupons. Also, consider the problem where we have c_1 Poisson processes of rate $p_1, \ldots, c_n$ Poisson processes of rate p_n, all independent, and wish to wait for all of them to occur at least once; let E be the time until this happens. Since the total rate of all these Poisson processes is 1, the time between any Poisson event and the subsequent event is exponentially distributed with parameter 1. Therefore, E has the same distribution as $Z_1 + \cdots + Z_D$, where $Z_1, Z_2, \ldots$ are independent exponential random variables with parameter 1, so

$$\mathbf{E}E = \mathbf{E}(Z_1 + \cdots + Z_D) = \mathbf{E}(\mathbf{E}(Z_1 + \cdots + Z_D | D)) = \mathbf{E}D \tag{5}$$

and

$$\begin{aligned} \mathbf{E}(E^2) &= \mathbf{E}((Z_1 + \cdots + Z_D)^2) \\ &= \mathbf{E}(\mathbf{E}((Z_1 + \cdots + Z_D)^2 | D)) \\ &= \mathbf{E}(D(D-1) + 2D) \\ &= \mathbf{E}(D(D+1)). \end{aligned} \tag{6}$$

The probability that all our Poisson processes have completed by time t is clearly $\prod_{1 \le i \le n} (1 - e^{-tp_i})^{c_i}$. Therefore, using (5), we can write

$$\mathbf{E}D = \mathbf{E}E = \int_0^\infty 1 - \prod_{1 \le i \le n} (1 - e^{-tp_i})^{c_i} \, dt, \tag{7}$$

and similarly, using (6),

$$\mathbf{E}(D(D+1)) = \mathbf{E}(E^2) = 2 \int_0^\infty t\alpha(t) \, dt, \tag{8}$$

where we have written

$$\alpha(t) = 1 - \prod_{1 \le i \le n} (1 - e^{-tp_i})^{c_i}.$$

Suppose now that

$$\sum_{1\le i\le n} \frac{c_i}{c_j^{p_i/p_j}} \le K, \qquad \text{where } K \ge 1, \tag{9}$$

and

$$\frac{p_i}{p_j} \ge \epsilon > 0, \qquad \text{for all } i = 1,\, \ldots,\, n. \tag{10}$$

In this case, we will have

$$0 \le p_j(\mathbf{E}D) - \log c_j \le \frac{K+1}{\epsilon} \tag{11}$$

and

$$\frac{2\epsilon}{e(\epsilon+1)(Ke)^{1/\epsilon}} \le p_j^2 \operatorname{Var}\ D \le \frac{2Ke\Gamma(\epsilon)}{\epsilon}. \tag{12}$$

To prove this, observe that substituting $y = c_j e^{-p_j t}$ into (7) yields

$$\begin{aligned} p_j \mathbf{E}D &= \int_0^{c_j} (1 - \prod_{1\le i\le n} (1 - (\frac{y}{c_j})^{p_i/p_j})^{c_i})\ \frac{dy}{y} \\ &= \int_0^{c_j} \psi(y)\ \frac{dy}{y}, \qquad \text{say.} \end{aligned}$$

Now, we always have $0 \le \psi(y) \le 1$, and if $y \le e^{-1/\epsilon}$, then $(y/c_j)^{p_i/p_j} \le y^{p_i/p_j} \le y^\epsilon \le e^{-1}$. As we have seen, $1 - z \ge e^{-ez}$ if $0 \le z \le e^{-1}$, so if $y \le e^{-1/\epsilon}$,

$$1 - (\frac{y}{c_j})^{p_i/p_j} \ge \exp(-e(\frac{y}{c_j})^{p_i/p_j})$$

and so

$$\begin{aligned} \psi(y) &\le 1 - \exp(-e \sum_{1\le i\le n} c_i (\frac{y}{c_j})^{p_i/p_j}) \\ &\le 1 - \exp(-e y^\epsilon \sum_{1\le i\le n} c_i c_j^{-p_i/p_j}) \\ &\le 1 - \exp(-eKy^\epsilon) \\ &\le eKy^\epsilon. \end{aligned}$$

Therefore,

$$\begin{aligned} p_j \mathbf{E}D &= \int_0^{c_j} \psi(y)\ \frac{dy}{y} \\ &\le \int_0^{e^{-1/\epsilon}} eKy^\epsilon\ \frac{dy}{y} + \int_{e^{-1/\epsilon}}^{c_j} \frac{dy}{y} \\ &= \frac{K+1}{\epsilon} + \log c_j, \end{aligned}$$

and we have the upper bound in (11). For the lower bound, observe that D is bounded below by the time to collect all c_j coupons of probability p_j. Given that we have already collected k of these coupons, the time to collect the next will be a geometric random variable with parameter $(c_j - k)p_j$, so the expected time to collect the next will be $p_j^{-1}/(c_j - k)$. Summing this over $k = 0, \ldots, c_j - 1$ gives us

$$\mathbf{E}D \geq p_j^{-1} \sum_{1 \leq k \leq c_j} \frac{1}{k} \geq p_j^{-1} \int_1^{c_j+1} \frac{dk}{k} = p_j^{-1} \log(c_j + 1) \geq p_j^{-1} \log c_j,$$

as desired. This proves (11). Now from (8),

$$\mathbf{E}(D(D+1)) = 2 \int_0^\infty t\alpha(t)\, dt = 2 \int_0^\infty \int_0^t \alpha(t)\, du\, dt$$

and from (7),

$$(\mathbf{E}D)^2 = \left(\int_0^\infty \alpha(t)\, dt \right)^2 = \int_0^\infty \int_0^\infty \alpha(t)\alpha(u)\, du\, dt = 2 \int_0^\infty \int_0^t \alpha(t)\alpha(u)\, du\, dt$$

so

$$\mathbf{E}(D(D+1)) - (\mathbf{E}D)^2 = 2 \int_0^\infty \int_0^t (1 - \alpha(u))\alpha(t)\, du\, dt.$$

Substituting $y = c_j e^{-p_j t}$ and $z = c_j e^{-p_j u}$ into this yields

$$p_j^2(\mathbf{E}(D(D+1)) - (\mathbf{E}D)^2) = 2 \int_0^{c_j} \int_y^{c_j} \frac{(1 - \psi(z))\psi(y)}{yz}\, dz\, dy.$$

Now as we have already seen, $\psi(y) \leq eKy^\epsilon$ if $y \leq e^{-1/\epsilon}$, and this is also true for $y \geq e^{-1/\epsilon}$ as $\psi(y) \leq 1$ and $K \geq 1$. Also, we have

$$1 - \psi(z) = \prod_{1 \leq i \leq n} (1 - (\frac{z}{c_j})^{p_i/p_j})^{c_i} \leq (1 - \frac{z}{c_j})^{c_j} \leq e^{-z},$$

so

$$\begin{aligned}
p_j^2(\mathbf{E}(D(D+1)) - (\mathbf{E}D)^2) &\leq 2 \int_0^{c_j} \int_y^{c_j} \frac{e^{1-z} K y^\epsilon}{yz}\, dz\, dy \\
&\leq 2 \int_0^\infty \int_y^\infty \frac{e^{1-z} K y^\epsilon}{yz}\, dz\, dy \\
&= 2Ke \int_0^\infty \frac{e^{-z}}{z} \int_0^z y^{\epsilon-1}\, dy\, dz \\
&= 2Ke \int_0^\infty \frac{e^{-z} z^{\epsilon-1}}{\epsilon}\, dz \\
&= \frac{2Ke\Gamma(\epsilon)}{\epsilon},
\end{aligned}$$

and we have the upper bound in (12). For the lower bound, observe that since $c_j \geq 1$,

$$\begin{aligned}
p_j^2(\mathbf{E}(D(D+1)) - (\mathbf{E}D)^2) &\geq 2\int_0^{(Ke)^{-1/\epsilon}} \int_y^{(Ke)^{-1/\epsilon}} \frac{(1-\psi(z))\psi(y)}{yz}\, dz\, dy \\
&\geq 2\int_0^{(Ke)^{-1/\epsilon}} \int_y^{(Ke)^{-1/\epsilon}} \frac{(1-e^{-y})(1-Kez^{\epsilon})}{yz}\, dz\, dy \\
&= 2\int_0^{(Ke)^{-1/\epsilon}} \int_0^{z} \frac{(1-e^{-y})(1-Kez^{\epsilon})}{yz}\, dy\, dz.
\end{aligned}$$

Now since $Ke \geq 1$ we must have $y \leq 1$. Therefore, $e^{-y} \leq 1 - y/e$ and $(1 - e^{-y})/y \geq e^{-1}$, so

$$\begin{aligned}
p_j^2(\mathbf{E}(D(D+1)) - (\mathbf{E}D)^2) &\geq \frac{2}{e}\int_0^{(Ke)^{-1/\epsilon}} 1 - Kez^{\epsilon} dz \\
&= \frac{2\epsilon}{e(\epsilon+1)(Ke)^{1/\epsilon}},
\end{aligned}$$

which is the lower bound in (12).

We now apply these results to the case where coupons are words. Let q_1, …, q_r be the q_i's of minimum probability, so that $\underline{q} = q_1 = \ldots = q_r$. We will identify the type of a word with the number of $r+1$s, …, Ms that it contains; if d_{r+1}, …, d_M are nonnegative integers satisfying $d_{r+1} + \cdots + d_M \leq L$, we will have a type of word containing d_{r+1} $r+1$s, …, d_M Ms, which will have $r^{L-d_{r+1}-\cdots-d_M}\binom{L}{d_{r+1}\ \ldots\ d_M\ L-d_{r+1}-\cdots-d_M}$ words, each of probability $\underline{q}^L \prod_{r+1\leq i\leq M}(q_i/\underline{q})^{d_i}$.

If $r > 1$, we take the type j to be the type where $d_{r+1} = \ldots = d_M = 0$. The coupons in this type have probability $\underline{q}^L$, which is the smallest probability of any coupon; therefore, (10) is satisfied with $\epsilon = 1$. For (9), observe that there are r^L coupons in this type, so we need to bound

$$\sum r^{L-d_{r+1}-\cdots-d_M}\binom{L}{d_{r+1}\ \ldots\ d_M\ L-d_{r+1}-\cdots-d_M}(r^L)^{-\prod_{r+1\leq i\leq M}(q_i/\underline{q})^{d_i}} = A_L, \qquad \text{say,}$$

where the sum is over all d_{r+1}, …, $d_M \geq 0$ with $d_{r+1} + \cdots + d_M \leq L$. However, if we let $\delta > 0$ be the minimum of $\log(q_i/\underline{q})$ over $i = r+1$, …, M, then

$$\prod_{r+1\leq i\leq M}(q_i/\underline{q})^{d_i} \geq \exp\delta(d_{r+1}+\cdots+d_M) \geq 1 + \delta(d_{r+1}+\cdots+d_M)$$

so

$$\begin{aligned}
A_L &\leq \sum_{d_{r+1},\ldots,d_M\geq 0} r^{L-d_{r+1}-\cdots-d_M} L^{d_{r+1}+\cdots+d_M} (\prod_{r+1\leq i\leq M} d_i!)^{-1} r^{-L(1+\delta(d_{r+1}+\cdots+d_M))} \\
&= \exp\left((M-r)Lr^{-(1+\delta L)}\right).
\end{aligned}$$

Now as $L \to \infty$, $Lr^{-(1+\delta L)} \to 0$, so A_L is bounded above uniformly in L. Plugging the coupon probability for this type and the number of coupons in this type into (11) and (12) now proves part 1 of this theorem.

If $r = 1$, let $\delta > 0$, as before, be the minimum of $\log(q_i/\underline{q})$ over $i = 2, \ldots, M$. Then

$$(d_2 + \cdots + d_M) \prod_{2 \le j \le M} (\underline{q}/q_j)^{d_j} \le (d_2 + \cdots + d_M) \exp(-\delta(d_2 + \cdots + d_M)),$$

and the right-hand side of this vanishes as $d_2 + \cdots + d_M \to \infty$; therefore, the supremum of the left-hand side must be finite and attained, as claimed. Suppose that it is attained at $d_2 = b_2, \ldots, d_M = b_M$. Since the left-hand side is zero if $d_2, \ldots, d_M$ are all zero, we cannot have $b_2, \ldots, b_M$ all zero. If $d_2, \ldots, d_M$ are not all zero, we then have

$$\begin{aligned}
\prod_{2 \le i \le M} (q_i/\underline{q})^{d_i - b_i} &= \prod_{2 \le i \le M} (q_i/\underline{q})^{d_i} \Bigg/ \prod_{2 \le i \le M} (q_i/\underline{q})^{b_i} \\
&= \frac{d_2 + \cdots + d_M}{b_2 + \cdots + b_M} \left(\frac{b_2 + \cdots + b_M}{\prod_{2 \le i \le M} (q_i/\underline{q})^{b_i}} \Bigg/ \frac{d_2 + \cdots + d_M}{\prod_{2 \le i \le M} (q_i/\underline{q})^{d_i}} \right) \\
&\ge \frac{d_2 + \cdots + d_M}{b_2 + \cdots + b_M}, \qquad (13)
\end{aligned}$$

and since this is also true if $d_2 = \ldots = d_M = 0$, (13) is always true. We now assume that $L > b_2 + \cdots + b_M$ and take the type j to be the type where $d_2 = b_2, \ldots, d_M = b_M$. The coupons in this type have probability $\underline{q}^L \prod_{2 \le i \le M} (q_i/\underline{q})^{b_i}$; since there are only a finite number of choices for $d_2, \ldots, d_M$ which yield a smaller coupon probability, and the ratio of these probabilities, $\prod_{2 \le i \le M} (q_i/\underline{q})^{d_i - b_i}$, depends only on the b_i's and d_i's and is independent of L, (10) is satisfied. For (9), we need to bound

$$\sum \binom{L}{d_2 \ \ldots \ d_M \ L - d_2 - \cdots - d_M} \binom{L}{b_2 \ \ldots \ b_M \ L - b_2 - \cdots - b_M}^{-\prod_{2 \le i \le M} (q_i/\underline{q})^{d_i - b_i}} = B_L,$$

say, where the sum is over all $d_2, \ldots, d_M \ge 0$ with $d_2 + \cdots + d_M \le L$. By (13), however, B_L is no more than

$$\sum_{d_2, \ldots, d_M \ge 0} \binom{L}{d_2 \ \ldots \ d_M \ L - d_2 - \cdots - d_M} \binom{L}{b_2 \ \ldots \ b_M \ L - b_2 - \cdots - b_M}^{-(d_2 + \cdots + d_M)/(b_2 + \cdots + b_M)}.$$

This is bounded above by

$$\sum_{d_2, \ldots, d_M \ge 0} L^{d_2 + \cdots + d_M} (\prod_{2 \le i \le M} d_i!)^{-1} \binom{L}{b_2 \ \ldots \ b_M \ L - b_2 - \cdots - b_M}^{-(d_2 + \cdots + d_M)/(b_2 + \cdots + b_M)},$$

which equals

$$\exp\left((M-1) L \binom{L}{b_2 \ \ldots \ b_M \ L - b_2 - \cdots - b_M}^{-1/(b_2 + \cdots + b_M)} \right).$$

However, $\binom{L}{b_2 \ \dots \ b_M \ L-b_2-\cdots-b_M}$ is bounded below by a positive constant, $\eta > 0$ say, times $(L - b_2 - \cdots - b_M)^{b_2+\cdots+b_M}$. It follows that

$$\sup_{L \geq b_2+\cdots+b_M+1} L \binom{L}{b_2 \ \dots \ b_M \ L - b_2 - \cdots - b_M}^{-1/(b_2+\cdots+b_M)} \leq \eta^{-1/(b_2+\cdots+b_M)}(b_2+\cdots+b_M+1).$$

This proves that $\sup_{L \geq b_2+\cdots+b_M+1} B_L < \infty$, which proves (9). Using (11) and (12) now gives us part 2 of the theorem. ∎