Binomial Randomization 3/13/02

Define $\mathbb{S}^n$ to be the product space $\{0,1,\ldots,m_1\} \times \cdots \times \{0,1,\ldots,m_n\}$ and let $\mathbb{S}^n_t$ be the set of all vectors $(s_1,\ldots,s_n)$ in $\mathbb{S}^n$ such that $s_1 + \ldots + s_n = t$.

Suppose that $(X_1,\ldots,X_n)$ is a **multivariate hypergeometric** random vector. That is, for all $(s_1,\ldots,s_n) \in \mathbb{S}^n_t$,

$$P(X_1 = s_1,\ldots,X_n = s_n) = \frac{\binom{m_1}{s_1}\cdots\binom{m_n}{s_n}}{\binom{M}{t}}$$

where $M = m_1 + \ldots + m_n$ and $t = s_1 + \ldots + s_n$.

For the purposes of Theorems 1 and 2 which follow we we will suppose that $Y_1,\ldots,Y_n$ are independent **binomial** random variables parameterized such that for $j = 1,2,\ldots,n$

$$P(Y_j = y) = \binom{m_j}{y}\left(\frac{\theta}{\theta+1}\right)^y\left(1-\frac{\theta}{\theta+1}\right)^{m_j-y} \qquad y = 0,1,\ldots,m_j.$$

It is well known that in this case $Y_1 + \ldots + Y_n$ will follow a binomial distribution with parameters M and θ. Therefore

$$P\left(Y_1 = s_1,\ldots,Y_n = s_n \middle| \sum_{i=1}^n \mathrm{Y}_i = t\right)$$

$$= \frac{P(Y_1 = s_1,\ldots,Y_n = s_n)}{P\left(\sum_{i=1}^n \mathrm{Y}_i = t\right)}\, \mathrm{I}_{\{t\}}\left(\sum_{i=1}^n s_i\right)$$

$$= \frac{\prod_{i=1}^n \binom{m_i}{s_i}\theta^{s_i}\left(\frac{1}{\theta+1}\right)^{m_i}}{\binom{M}{t}\theta^t\left(\frac{1}{\theta+1}\right)^M}\, \mathrm{I}_{\{t\}}\left(\sum_{i=1}^n s_i\right)$$

$$= \frac{\frac{\binom{m_1}{s_1}\cdots\binom{m_n}{s_n}}{\binom{M}{t}}\binom{M}{t}\,\theta^{\sum_{i=1}^{n}s_i}\left(\frac{1}{\theta+1}\right)^M}{\binom{M}{t}\theta^t\left(\frac{1}{\theta+1}\right)^M}\ \mathrm{I}_{\{t\}}\left(\sum_{i=1}^{n}s_i\right)$$

$$= P(X_1 = s_1,\ldots,X_n = s_n)\ \mathrm{I}_{\{t\}}\left(\sum_{i=1}^{n}s_i\right)$$

Theorem 1.

$$E_t(g(X_1,\ldots,X_n)) = \frac{1}{\binom{M}{t}t!}\frac{d^t}{d\theta^t}\left((1+\theta)^M E(g(Y_1,\ldots,Y_n))\right)\Big|_{\theta=0}$$

where the t in $E_t(\,\cdot\,)$ is used to denote that $X_1 + \ldots + X_n = t$.

Proof.

$$E(g(Y_1,\ldots,Y_n)) = \sum_{t=0}^{M} E\left(g(Y_1,\ldots,Y_n) \mid \sum_{i=1}^{n}\mathrm{Y}_i = t\right)P\left(\sum_{i=1}^{n}\mathrm{Y}_i = t\right)$$

$$= \sum_{t=0}^{M}\sum_{\mathbb{S}_t^n} g(s_1,\ldots,s_n)P\left(Y_1 = s_1,\ldots,Y_n = s_n \mid \sum_{i=1}^{n}\mathrm{Y}_i = t\right)P\left(\sum_{i=1}^{n}\mathrm{Y}_i = t\right)$$

$$= \sum_{t=0}^{M}\sum_{\mathbb{S}_t^n} g(s_1,\ldots,s_n)P(X_1 = s_1,\ldots,X_n = s_n)\ P\left(\sum_{i=1}^{n}\mathrm{Y}_i = t\right)$$

$$= \sum_{t=0}^{M} E_t(g(X_1,\ldots,X_n))\ P\left(\sum_{i=1}^{n}\mathrm{Y}_i = t\right)$$

$$= \sum_{t=0}^{M} E_t(g(X_1,\ldots,X_n))\binom{M}{t}\theta^t\left(\frac{1}{\theta+1}\right)^M$$

and

$$(1+\theta)^M E(g(Y_1,\ldots,Y_n)) = \sum_{t=0}^{M}E_t(g(X_1,\ldots,X_n))\binom{M}{t}\theta^t$$

Therefore,

$$\frac{d^r}{d\theta^r}\left((1+\theta)^M E(g(Y_1,\ldots,Y_n))\right)\bigg|_{\theta=0} = \frac{d^r}{d\theta^r}\left(\sum_{t=0}^{M} E_t(g(X_1,\ldots,X_n))\binom{M}{t}\theta^t\right)\bigg|_{\theta=0}$$

$$= \sum_{t=0}^{M} E_t(g(X_1,\ldots,X_n))\binom{M}{t}\left(\frac{d^r}{d\theta^r}\theta^t\right)\bigg|_{\theta=0}$$

$$= \sum_{t=0}^{M} E_t(g(X_1,\ldots,X_n))\binom{M}{t} r!\,\mathrm{I}_{\{r\}}(t)$$

$$= E_r(g(X_1,\ldots,X_n))\binom{M}{r} r!\,\mathrm{I}_{\{0,1,\ldots,M\}}(r)$$

Hence,

$$E_t(g(X_1,\ldots,X_n)) = \frac{1}{\binom{M}{t}t!}\frac{d^t}{d\theta^t}\left((1+\theta)^M E(g(Y_1,\ldots,Y_n))\right)\bigg|_{\theta=0}$$

Theorem 2.

Let $\mathcal{A} \subset \mathbb{S}^n$ and define $\mathcal{A}_t = \mathcal{A} \cap \mathbb{S}^n_t$. Then for $t \geq 0$,

$$P((X_1,\ldots,X_n) \in \mathcal{A}_t) = \frac{1}{\binom{M}{t}t!}\frac{d^t}{d\theta^t}\left((1+\theta)^M P((Y_1,\ldots,Y_n) \in \mathcal{A})\right)\bigg|_{\theta=0}$$

Proof.

Apply Theorem 1 with

$$g(s_1,\ldots,s_n) = \begin{cases} 1 & (s_1,\ldots,s_n) \in \mathcal{A} \\ 0 & \text{else} \end{cases}$$

so that

$$E_t(g(X_1,\ldots,X_n)) = P((X_1,\ldots,X_n) \in \mathcal{A}_t)$$

and

$$E(g(Y_1,\ldots,Y_n)) = P((Y_1,\ldots,Y_n) \in \mathcal{A}). \qquad \square$$

For the purposes of Theorem 3 which follows we will suppose that $Y_1,\ldots,Y_n$ are independent **binomial** random variables parameterized such that for $j = 1,2,\ldots,n$

$$P(Y_j = y) = \binom{m_j}{y} p^y (1-p)^{m_j - y} \qquad y = 0, 1, \ldots, m_j.$$

Suppose an urn contains m_1 objects of Type $1,\ldots,m_n$ objects of Type n and that objects are drawn from this urn without replacement. Let $M = m_1 + \ldots + m_n$. When q_i objects of Type i have been drawn we will say Type i has reached its ***quota***.

Let $W_{(r:q_1,\ldots,q_n)}(m_1,\ldots,m_n;n) \equiv W_{r:Q}$ represent the waiting time until exactly r different types have reached their quota.

Let $E\left(W_{r:Q}^{[k]}\right)$ represent the k^{th} ascending moment of $W_{r:Q}$. That is,

$$E\left(W_{r:Q}^{[k]}\right) = E((W_{r:Q}+0)(W_{r:Q}+1)\cdots(W_{r:Q}+k-1))$$

<u>Theorem 3.</u>

$$E\left(W_{r:Q}^{[k]}\right) = \frac{k(M+k)!}{M!}\left(\int_0^1 p^{k-1}\, P((Y_1,\ldots,Y_n) \in \mathbb{A}_{Q:r})\mathrm{d}p\right)$$

and

$$E\left(W_{r:Q}^{[k]} - W_{r-1:Q}^{[k]}\right) = \frac{k(M+k)!}{M!}\left(\int_0^1 p^{k-1}\, P((Y_1,\ldots,Y_n) \in \mathbb{B}_{Q:r})\mathrm{d}p\right)$$

where

(1) $\mathbb{A}_{Q:r}$ is the event that at least $n-r+1$ of the (independent) events $\mathcal{A}_1,\ldots,\mathcal{A}_n$ occur

$\mathbb{B}_{Q:r}$ is the event that exactly $n-r+1$ of the (independent) events $\mathcal{A}_1,\ldots,\mathcal{A}_n$ occur

(2) $\mathcal{A}_j$ is the event that $Y_j < q_j$.

<u>Proof</u>

Define $N_{(q_1,\ldots,q_n)}(t) \equiv N_Q(t)$ to be the number of Types that have <u>not</u> reached their quota after t balls have been drawn out.

Define N_Q^B as the number of events amongst $\mathcal{A}_1,\ldots,\mathcal{A}_n$ that occur.

It follows that

$$W_{r:Q} > t \Leftrightarrow N_Q(t) > n - r$$

However, it follows from

$$E(g(Y_1,\ldots,Y_n)) = \sum_{t=0}^{M} E\left(g(Y_1,\ldots,Y_n) \mid \sum_{i=1}^{n} \mathrm{Y}_i = t\right) P\left(\sum_{i=1}^{n} \mathrm{Y}_i = t\right)$$

that

$$P\left(N_Q^B > n - r\right) = \sum_{t=0}^{M} P(N_Q(t) > n - r)\binom{M}{t} p^t (1-p)^{M-t}.$$

Thus,

$$\begin{aligned}
&\int_0^1 p^{k-1}\left(P\left(N_Q^B > n - r\right)\right)\mathrm{d}p \\
&= \int_0^1 p^{k-1}\left(\sum_{t=0}^{M} P(N_Q(t) > n - r)\binom{M}{t} p^t (1-p)^{M-t}\right)\mathrm{d}p \\
&= \int_0^1 p^{k-1}\left(\sum_{t=0}^{M} P(W_{r:Q} > t)\binom{M}{t} p^t (1-p)^{M-t}\right)\mathrm{d}p \\
&= \sum_{t=0}^{M} P(W_{r:Q} > t)\binom{M}{t}\left(\int_0^1 p^{k+t-1}(1-p)^{M-t}\mathrm{d}p\right) \\
&= \sum_{t=0}^{M} P(W_{r:Q} > t)\binom{M}{t}\left(\frac{(M-t)!}{\prod_{j=0}^{M-t}(j+k+t)}\right) \\
&= \frac{M!}{(M+k)!}\sum_{t=0}^{M} P(W_{r:Q} > t)\frac{(t+k-1)!}{t!}
\end{aligned}$$

$$\begin{aligned}
&= \frac{M!}{(M+k)!} \sum_{t=0}^{M} P(W_{r:Q} > t)(t+k-1)_{[k-1]} \\
&= \frac{M!}{(M+k)!} \sum_{t=k-1}^{M+k-1} P(W_{r:Q} + k - 1 > t)\ t_{[k-1]} \\
&= \frac{M!}{(M+k)!} \frac{1}{k} E\Big((W_{r:Q} + k - 1)_{[k]}\Big)
\end{aligned}$$

But

$$(W_{r:Q} + k - 1)_{[k]} \ \equiv \ W_{r:Q}^{[k]}$$

hence

$$\int_0^1 p^{k-1}\big(P\big(N_Q^B > n - r\big)\big)\mathrm{d}p = \frac{M!}{(M+k)!} \frac{1}{k} E\Big(W_{r:Q}^{[k]}\Big)$$

and

$$E\Big(W_{r:Q}^{[k]}\Big) = \frac{k\ (M+k)!}{M!} \int_0^1 p^{k-1}\big(P\big(N_Q^B > n - r\big)\big)\mathrm{d}p$$

where

$$P\big(N_Q^B > n - r\big) = P(\text{at least } n - r + 1 \text{ types do \underline{not} obtain their quota}|\text{Binomial model})$$

Therefore,

$$E\Big(W_{r:Q}^{[k]} - W_{r-1:Q}^{[k]}\Big) = \frac{k\ (M+k)!}{M!} \int_0^1 p^{k-1}\big(P\big(N_Q^B = n - r + 1\big)\big)\mathrm{d}p$$

where

$$P\big(N_Q^B = n - r + 1\big) = P(\text{exactly } n - r + 1 \text{ types do \underline{not} obtain their quota}|\text{Binomial model}). \qquad \square$$

Theorem 4.

Suppose we draw t balls without replacement from an urn initially containing m_j balls of color j, $j = 1, \ldots, n$.

Again letting C_j equal the number of times color j is selected, now let D_k equal the number of C_j's which equal k, $k \in \{0, 1, \ldots, t\}$.

If $m_1 = \ldots = m_n = m$, then

$$\mathrm{E}(\Psi(D_0, D_1, \ldots, D_t, 0, 0, \ldots)) = \frac{1}{\binom{mn}{t} t!} \frac{d^t}{d\lambda^t} \frac{d^n}{d\theta^n} \left(e^{\theta(1+\lambda)^m} \mathrm{E}(\Psi(Z_0, Z_1, \ldots)) \right) \Big|_{\substack{\lambda=0 \\ \theta=0}} \tag{0.0.1}$$

where $Z_0, Z_1, \ldots$ are independent and $Z_j \sim \text{Poisson}\left(\binom{m}{j} \theta \lambda^j\right)$.

Models:

Multivariate Hypergeometric $\equiv$ Grouped Fermi Dirac Distribution, Urns with cells, at most one ball per cell, balls identical.

Applications

Problem 1.

Suppose we draw t balls without replacement from an urn containing m copies of each of n different colored balls. Let V_r equal the number of colors which are selected exactly r times. Find $P(V_r = v)$.

Answer

$$\frac{\binom{n}{v}\binom{m}{r}^v}{\binom{mn}{t}} \sum_{j=0}^{n-v} (-1)^j \binom{mn - m(j+v)}{t - r(j+v)} \binom{n-v}{j} \binom{m}{r}^j$$

provided $t \in \{rv, \ldots, rn\}$.

Proof

Define

$$\mathcal{A}_t = \left\{ a_1,\ldots,a_n \middle| \sum_{j=0}^{n} \mathrm{I}_{\{r\}}(a_j) \ = v \text{ and } a_1 + \ldots + a_n = t \right\}$$
$$\mathcal{A} = \left\{ a_1,\ldots,a_n \middle| \sum_{j=0}^{n} \mathrm{I}_{\{r\}}(a_j) \ = v \right\}$$

It follows from Theorem 2 that

$$P(V_r = v) \ = \ P((X_1,\ldots,X_n) \in \mathcal{A}_t)$$
$$= \frac{1}{\binom{mn}{t} t!} \frac{d^t}{d\theta^t} ((1+\theta)^{mn} P((Y_1,\ldots,Y_n) \in \mathcal{A})) \bigg|_{\theta=0}$$

Let B_j represent the event that $Y_j = r$. Then by the generalized inclusion-exclusion principle we have

$$P((Y_1,\ldots,Y_n) \in \mathcal{A})$$

$$= P(\text{exactly } v \text{ of the events } B_1, \ldots, B_n \text{ occur})$$

$$= \sum_{u=0}^{n-v} (-1)^u \binom{u+v}{v}\binom{n}{u+v} \left(P(Y_1 = r)\right)^{u+v}$$

$$= \sum_{u=0}^{n-v} (-1)^u \binom{u+v}{v}\binom{n}{u+v} \left(\binom{m}{r}\left(\frac{\theta}{\theta+1}\right)^r\left(1-\frac{\theta}{\theta+1}\right)^{m-r}\right)^{u+v}$$

$$= \sum_{u=0}^{n-v} (-1)^u \binom{n}{v}\binom{n-v}{u} \binom{m}{r}^{u+v} \theta^{r(u+v)}(1+\theta)^{-m(u+v)}$$

Therefore,

$$P(V_r = v)$$

$$= \frac{1}{\binom{mn}{t}t!}\frac{d^t}{d\theta^t}\left((1+\theta)^{mn}\binom{n}{v}\binom{m}{r}^v \sum_{u=0}^{n-v} (-1)^u \binom{n-v}{u}\binom{m}{r}^u \theta^{r(u+v)}(1+\theta)^{-m(u+v)}\right)\Bigg|_{\theta=0}$$

$$= \frac{1}{\binom{mn}{t}t!}\frac{d^t}{d\theta^t}\left(\binom{n}{v}\binom{m}{r}^v \sum_{u=0}^{n-v} (-1)^u \binom{n-v}{u}\binom{m}{r}^u \theta^{r(u+v)}(1+\theta)^{m(n-u-v)}\right)\Bigg|_{\theta=0}$$

$$= \frac{1}{\binom{mn}{t}t!}\frac{d^t}{d\theta^t}\left(\binom{n}{v}\binom{m}{r}^v \sum_{u=0}^{n-v} \sum_{j=0}^{m(n-u-v)} (-1)^u \binom{m(n-u-v)}{j}\binom{n-v}{u}\binom{m}{r}^u \theta^{r(u+v)}\theta^j\right)\Bigg|_{\theta=0}$$

$$= \frac{1}{\binom{mn}{t}t!}\binom{n}{v}\binom{m}{r}^v \sum_{u=0}^{n-v} \sum_{j=0}^{m(n-u-v)} (-1)^u \binom{m(n-u-v)}{j}\binom{n-v}{u}\binom{m}{r}^u \left(\frac{d^t}{d\theta^t}\theta^{r(u+v)+j}\right)\Bigg|_{\theta=0}$$

$$= \frac{1}{\binom{mn}{t}t!}\binom{n}{v}\binom{m}{r}^v \sum_{u=0}^{n-v} \sum_{j=0}^{m(n-u-v)} (-1)^u \binom{m(n-u-v)}{j}\binom{n-v}{u}\binom{m}{r}^u t!\, \mathrm{I}_{\{r(u+v)+j\}}(t)$$

$$= \frac{1}{\binom{mn}{t} t!} \binom{n}{v} \binom{m}{r}^v \sum_{u=0}^{n-v} \sum_{j=0}^{m(n-u-v)} (-1)^u \binom{m(n-u-v)}{j} \binom{n-v}{u} \binom{m}{r}^u t!\, \mathrm{I}_{\{t-r(u+v)\}}(j)$$

$$= \frac{1}{\binom{mn}{t}} \binom{n}{v} \sum_{u=0}^{n-v} (-1)^u \binom{m(n-u-v)}{t-r(u+v)} \binom{n-v}{u} \binom{m}{r}^{v+u} \; \mathrm{I}_{\{0,\dots,m(n-u-v)\}}(t-r(u+v))$$

$$= \frac{\binom{n}{v}}{\binom{mn}{t}} \sum_{u=0}^{n-v} (-1)^u \binom{m(n-u-v)}{t-r(u+v)} \binom{n-v}{u} \binom{m}{r}^{v+u} \qquad \text{provided } t \in \{rv, \dots, rn\}$$

$$= \frac{\binom{n}{v}\binom{m}{r}^v}{\binom{mn}{t}} \sum_{u=0}^{n-v} (-1)^u \binom{mn-m(u+v)}{t-r(u+v)} \binom{n-v}{u} \binom{m}{r}^u \qquad \text{provided } t \in \{rv, \dots, rn\} \quad \square$$

Alternative proof using Theorem 4.

Let $Z_r \sim \text{Poisson}\left(\binom{m}{r}\theta\lambda^r\right)$. Then

$$P(D_r = v) = \frac{1}{\binom{mn}{t} t!} \frac{d^t}{d\lambda^t} \frac{d^n}{d\theta^n} \left(e^{\theta(1+\lambda)^m} P(Z_r = v) \right) \Big|_{\substack{\lambda=0 \\ \theta=0}}$$

$$= \frac{1}{\binom{mn}{t} t!} \frac{d^t}{d\lambda^t} \frac{d^n}{d\theta^n} \left(e^{\theta(1+\lambda)^m} \frac{e^{-\binom{m}{r}\theta\lambda^r} \left(\binom{m}{r}\theta\lambda^r\right)^v}{v!} \right) \Big|_{\substack{\lambda=0 \\ \theta=0}}$$

$$= \frac{\binom{m}{r}^v}{\binom{mn}{t} t! v!} \frac{d^t}{d\lambda^t} \frac{d^n}{d\theta^n} \left(e^{\theta(1+\lambda)^m} e^{-\binom{m}{r}\theta\lambda^r} \theta^v \lambda^{rv} \right) \Big|_{\substack{\lambda=0 \\ \theta=0}}$$

$$= \frac{\binom{m}{r}^v}{\binom{mn}{t} t! v!} \frac{d^t}{d\lambda^t} \frac{d^n}{d\theta^n} \left(e^{\theta\left((1+\lambda)^m - \binom{m}{r}\lambda^r\right)} \theta^v \lambda^{rv} \right) \Big|_{\substack{\lambda=0 \\ \theta=0}}$$

$$= \frac{\binom{m}{r}^v}{\binom{mn}{t} t! v!} \frac{n!}{(n-v)!} \frac{d^t}{d\lambda^t} \left(\left((1+\lambda)^m - \binom{m}{r}\lambda^r\right)^{n-v} \lambda^{rv} \right) \Big|_{\lambda=0}$$

$$= \frac{\binom{m}{r}^v}{\binom{mn}{t} t! v!} \frac{n!}{(n-v)!} \frac{d^t}{d\lambda^t} \left(\sum_{j=0}^{n-v} (-1)^j \binom{n-v}{j} \binom{m}{r}^j (1+\lambda)^{m(n-v-j)} \lambda^{r(j+v)} \right) \Big|_{\lambda=0}$$

$$= \frac{\binom{m}{r}^v \binom{n}{v}}{\binom{mn}{t}} \sum_{j=0}^{n-v} (-1)^j \binom{n-v}{j} \binom{m}{r}^j \binom{m(n-v-j)}{t-r(j+v)}$$

EXTRA

Using Theorem 4,

Let $Z_j \sim$ Poisson$\left(\binom{m}{j}\theta\lambda^j\right)$. Then

$$P(D_0 = 0, D_1 = 0, \ldots, D_r = 0) = \frac{1}{\binom{mn}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\left(e^{\theta(1+\lambda)^m}P(V_0 = 0, V_1 = 0, \ldots, V_r = 0)\right)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{1}{\binom{mn}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\left(e^{\theta(1+\lambda)^m}P(V_0 = 0, V_1 = 0, \ldots, V_r = 0)\right)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{1}{\binom{mn}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\left(e^{\theta(1+\lambda)^m}\prod_{j=0}^{r}e^{-\binom{m}{j}\theta\lambda^j}\right)\Bigg|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{1}{\binom{mn}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\left(e^{\theta\left((1+\lambda)^m - \sum_{j=0}^{r}\binom{m}{j}\lambda^j\right)}\right)\Bigg|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{1}{\binom{mn}{t}t!}\frac{d^t}{d\lambda^t}\left(\left((1+\lambda)^m - \sum_{j=0}^{r}\binom{m}{j}\lambda^j\right)^n\right)\Bigg|_{\lambda=0}$$

$$= \frac{1}{\binom{mn}{t}t!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{d^t}{d\lambda^t}\left((1+\lambda)^{m(n-k)}\left(\sum_{j=0}^{r}\binom{m}{j}\lambda^j\right)^k\right)\Bigg|_{\lambda=0}$$

$$= \frac{1}{\binom{mn}{t}t!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}\frac{d^t}{d\lambda^t}\left((1+\lambda)^{m(n-k)}\left(\sum_{s=0}^{rk}C^*(m,s,r)\lambda^s\right)\right)\Bigg|_{\lambda=0}$$

where $C^*(m,s,r) = \sum_{\substack{(k_1,\ldots,k_r)\ni \\ k_1+\ldots+k_r=s \\ k_j\in\{0,1,\ldots,r\}}} \cdots \sum \binom{m}{k_1}\cdots\binom{m}{k_r}$

$$= \frac{1}{\binom{mn}{t}t!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}\sum_{s=0}^{rk}C^*(m,s,r)\frac{d^t}{d\lambda^t}\left((1+\lambda)^{m(n-k)}\lambda^s\right)\Big|_{\lambda=0}$$

$$= \frac{1}{\binom{mn}{t}}\sum_{k=0}^{n}(-1)^k\binom{n}{k}\sum_{s=0}^{rk}C^*(m,s,r)\binom{m(n-k)}{t-s}$$

$$= \frac{1}{\binom{mn}{t}}\sum_{k=0}^{n}\sum_{s=0}^{rk}(-1)^k\binom{n}{k}\binom{m(n-k)}{t-s}C^*(m,s,r)$$

Is there an expression for the C^* ???

Problem 2.

Suppose we draw t balls without replacement from an urn containing m copies of each of n different colored balls. Find the probability that every color is drawn at least once.

Answer

$$\frac{n!}{(nm)_t}\mathrm{C}(t,n,m)$$

where $\mathrm{C}(t,n,m)$ are the C-numbers defined by

$$\mathrm{C}(t,n,m) = \frac{1}{n!}\sum_{j=0}^{n}(-1)^{n-j}\binom{n}{j}(mj)_t.$$

and where we adopt the standard falling factorial notation

$$(a)_t = a(a-1)\cdots(a-t+1) = t!\binom{a}{t}.$$

Proof

The event that every color is drawn at least once is the event that $V_0 = 0$ using the notation of Problem 1. Therefore, from Problem 1,

$$\begin{aligned} P(V_0 = 0) &= \frac{\binom{n}{0}\binom{m}{0}^0}{\binom{mn}{t}} \sum_{i=0}^{n-0} (-1)^i \binom{mn - m(i+0)}{t - 0(i+0)} \binom{n-0}{i} \binom{m}{0}^i \\ &= \frac{1}{\binom{mn}{t}} \sum_{i=0}^{n} (-1)^i \binom{m(n-i)}{t} \binom{n}{i} \\ &= \frac{1}{\binom{mn}{t}} \sum_{j=0}^{n} (-1)^{n-j} \binom{mj}{t} \binom{n}{j} \qquad \text{letting } j = n - i \\ &= \frac{n!}{(nm)_t} \mathrm{C}(t,n,m). \qquad \square \end{aligned}$$

Continuing with this notation we can identify new numbers which we will name the **extended C-numbers** and define by

$$\mathrm{C}(t,n,m,r,v) = \frac{1}{n!} \sum_{j=0}^{n-v} (-1)^{n-v-j} \binom{n-v}{j} (mj)_{t-r(n-j)} \, (t)_{r(n-j)} \binom{n}{v} \binom{m}{r}^{n-j}$$

with the properties that

$$\mathrm{C}(t,n,m,r=0,v=0) = \mathrm{C}(t,n,m)$$

and

$$P(V_r = v) = \frac{n!}{(nm)_t} \mathrm{C}(t,n,m,r,v).$$

Problem 3.

Suppose we continue to draw balls without replacement from an urn containing m copies of each of n different colored balls until all n colors have been selected at least k times each. Let W_k equal the required waiting time (number of draws). Find the probability that it takes w draws to get secure one of every color, i.e. $P(W_1 = w)$. Then find the probability that it takes w draws to secure two of every color, i.e. $P(W_2 = w)$.+

Answer

Proof

Continuing with the notation leading up to Theorem 3, we let

$$W_k = W_{(n:k,\ldots,k)}(m,\ldots,m;n)$$

and similarly

$$N_k(t) = N_{(k,\ldots,k)}(t).$$

It follows that

$$W_k \le w \Leftrightarrow N_k(w) = 0$$

and

$$\begin{aligned} P(W_k = w) &= P(W_k \le w) - P(W_k \le w-1) \\ &= P(N_k(w) = 0) - P(N_k(w-1) = 0) \end{aligned}$$

Now consider first the case $k = 1$.

By Problem 2,

$$\begin{aligned} P(W_1 = w) &= \frac{n!}{(nm)_w}\mathrm{C}(w,n,m) - \frac{n!}{(nm)_{w-1}}\mathrm{C}(w-1,n,m) \\ &= \frac{n!}{(nm)_w}(\mathrm{C}(w,n,m) - (nm-w+1)\mathrm{C}(w-1,n,m)) \\ &= \frac{n!m}{(nm)_w}\mathrm{C}(w-1,n-1,m) \end{aligned}$$

This last equality can be verified directly. We see that

$$
\begin{aligned}
&\mathrm{C}(w,n,m)-(nm-w+1)\mathrm{C}(w-1,n,m)\\
&\quad=\left(\frac{1}{n!}\sum_{j=0}^{n}(-1)^{n-j}\binom{n}{j}(mj)_w\right)-(nm-w+1)\left(\frac{1}{n!}\sum_{j=0}^{n}(-1)^{n-j}\binom{n}{j}(mj)_{w-1}\right)\\
&\quad=\frac{1}{n!}\sum_{j=0}^{n}(-1)^{n-j}\binom{n}{j}((mj)_w-(nm-w+1)(mj)_{w-1})\\
&\quad=\frac{1}{n!}\sum_{j=0}^{n}(-1)^{n-j}\binom{n}{j}(m(j-n)(mj)_{w-1})\\
&\quad=m\left(\frac{1}{(n-1)!}\sum_{j=0}^{n-1}(-1)^{(n-1)-j}\binom{n-1}{j}(mj)_{w-1}\right)\\
&\quad=m\mathrm{C}(w-1,n-1,m)
\end{aligned}
$$

Now consider first the case $k=2$.

$$
\begin{aligned}
P(W_2=w)&=P(N_2(w)=0)-P(N_2(w-1)=0)\\
&=\frac{h_2(w,n,m)}{\binom{mn}{w}}-\frac{h_2(w-1,n,m)}{\binom{mn}{w-1}}
\end{aligned}
$$

where we define

$$
h_k(w,n,m)=\sum_{\substack{x_1+\ldots+x_n=w\\ x_i\in\{k,\ldots,m\},\,i\in\{1,\ldots,n\}}}\cdots\sum\binom{m}{x_1}\cdots\binom{m}{x_n}.
$$

Now take any set $\mathbb{A}\subset\{(x_1,\ldots,x_n)|\ x_i\in\{0,1,\ldots\},\ i\in\{1,\ldots,n\}\}$ and define

$$
V(\mathbb{A})=\sum_{\mathbb{A}}\binom{m}{x_1}\cdots\binom{m}{x_n}
$$

Then

$$
h_2(w,n,m)=V(\mathbb{D}_2)
$$

with

$$
\mathbb{D}_2=\{(x_1,\ldots,x_n)|\ x_1+\ldots+x_n=w,\ x_i\in\{2,\ldots,m\},\ i\in\{1,\ldots,n\}\}.
$$

Now define sets

$$\mathcal{B}_j = \left\{(x_1, \ldots, x_n) \middle|\, x_1 + \ldots + x_n = w,\, x_j = 1,\, x_i \in \{2, \ldots, m\}, i \in \{1, \ldots, n\},\, i \neq j\right\}$$

as subsets of universal set $\mathbb{D}_1$. Then,

$$\mathbb{D}_2 = \mathcal{B}_1' \cap \cdots \cap \mathcal{B}_n'$$

and

$$h_2(w, n, m) = V(\mathbb{D}_2)$$

$$= V(\mathcal{B}_1' \cap \cdots \cap \mathcal{B}_n')$$

$$= V(\mathbb{D}_1) - V(\mathcal{B}_1 \cup \cdots \cup \mathcal{B}_n)$$

$$= V(\mathbb{D}_1) - \sum_{j=0}^{n} (-1)^j \binom{n}{j} V(\mathcal{B}_1 \cap \cdots \cap \mathcal{B}_j)$$

$$= h_1(w, n, m) - \sum_{j=0}^{n} (-1)^j \binom{n}{j} \binom{m}{1}^j h_1(w - j, n - j, m)$$

$$= \sum_{j=1}^{n} (-1)^{j-1} \binom{n}{j} \binom{m}{1}^j h_1(w - j, n - j, m)$$

Therefore,

$$P(W_2 = w) = \frac{h_2(w,n,m)}{\binom{mn}{w}} - \frac{h_2(w-1,n,m)}{\binom{mn}{w-1}}$$

$$= \frac{\sum_{j=1}^{n}(-1)^{j-1}\binom{n}{j}\binom{m}{1}^j\, h_1(w-j,n-j,m)}{\binom{mn}{w}}$$

$$- \frac{\sum_{j=1}^{n}(-1)^{j-1}\binom{n}{j}\binom{m}{1}^j\, h_1(w-j-1,n-j,m)}{\binom{mn}{w-1}}$$

$$= \frac{\sum_{j=1}^{n}(-1)^{j-1}\binom{n}{j}\binom{m}{1}^j\, \frac{(n-j)!}{(w-j)!}\mathrm{C}(w-j,n-j,m)}{\binom{mn}{w}}$$

$$- \frac{\sum_{j=1}^{n}(-1)^{j-1}\binom{n}{j}\binom{m}{1}^j\, \frac{(n-j)!}{(w-j-1)!}\mathrm{C}(w-j-1,n-j,m)}{\binom{mn}{w-1}}$$

Problem 4.

Suppose we continue to draw balls without replacement from an urn containing m copies of each of n different colored balls until all n colors have been selected at least k times each. Let W_k equal the required waiting time (number of draws). Find $E(W_1)$ and $E(W_2)$.

Answer

$$\mathrm{E}\left(W^{[k]}_{r:(1,\ldots,1)}\right) = \frac{(mn+k)!}{(mn)!}\sum_{j=1}^{r}(-1)^{j-1}\frac{\binom{n-r+j-1}{n-r}\binom{n}{n-r+j}}{\binom{m(n-r+j)+k}{k}}$$

Proof

We have from Theorem 3 that

$$E\left(W^{[k]}_{r:Q}\right) = \frac{k(M+k)!}{M!}\left(\int_0^1 p^{k-1}\, P((Y_1,\ldots,Y_n) \in \mathbb{A}_{Q:r})\mathrm{d}p\right)$$

where

(1) $\mathbb{A}_{Q:r}$ is the event that at least $n-r+1$ of the (independent) events $\mathcal{A}_1,\ldots,\mathcal{A}_n$ occur

(2) $\mathcal{A}_j$ is the event that $Y_j < q_j$.

We consider the two special cases, $(r=1, q=1)$ and $(r=1, q=2)$. In this first case we will find the k^{th}ascending moment while in the later case we will consider only the first ascending moment.

If we take $q_1 = \ldots = q_n = q$, then by the general inclusion-exclusion principle,

$$P((Y_1,\ldots,Y_n) \in \mathbb{A}_{Q:r}) = P(\text{at least } n-r+1 \text{ of the events } \mathcal{A}_1,\ldots,\mathcal{A}_n \text{ occur})$$

$$= P(\text{at least } n-r+1 \text{ of the } Y_j < q\ )$$

$$= \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \mathbb{S}_{n-r+j+1}$$

$$= \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \binom{n}{n-r+j+1} (P(Y_j < q))^{n-r+j+1}$$

$$= \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \binom{n}{n-r+j+1} \left(\sum_{y=0}^{q-1} \binom{m}{y} p^y (1-p)^{m-y} \right)^{n-r+j+1}$$

$$= \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \binom{n}{n-r+j+1} \left(\sum_{y=0}^{q-1} \binom{m}{y} p^y (1-p)^{m-y} \right)^{n-r+j+1}$$

Therefore,

$$\int_0^1 p^{k-1}\ P((Y_1,\ldots,Y_n) \in \mathbb{A}_{Q:r}) \mathrm{d}p$$

$$= \int_0^1 p^{k-1} \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \binom{n}{n-r+j+1} \left(\sum_{y=0}^{q-1} \binom{m}{y} p^y (1-p)^{m-y} \right)^{n-r+j+1} \mathrm{d}p$$

Now suppose $q = 1$. Then

$$E\left(W^{[k]}_{r:(1,\ldots,1)}\right) = \frac{k(M+k)!}{M!}\left(\int_0^1 p^{k-1}\ P((Y_1,\ldots,Y_n)\in\mathbb{A}_{Q:r})\mathrm{d}p\right)$$

$$= \frac{k(M+k)!}{M!}\left(\int_0^1 p^{k-1}\sum_{j=0}^{r-1}(-1)^j\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}\left(\binom{m}{0}p^0(1-p)^{m-0}\right)^{n-r+j+1}\mathrm{d}p\right)$$

$$= \frac{k(M+k)!}{M!}\ \sum_{j=0}^{r-1}(-1)^j\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}\left(\int_0^1 p^{k-1}(1-p)^{m(n-r+j+1)}\mathrm{d}p\right)$$

$$= \frac{k(M+k)!}{M!}\ \sum_{j=0}^{r-1}(-1)^j\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}\left(\frac{(k-1)!(m(n-r+j+1))!}{(k+m(n-r+j+1))!}\right)$$

$$= \frac{(mn+k)!}{(mn)!}\ \sum_{j=0}^{r-1}(-1)^j\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}\left(\frac{k!(m(n-r+j+1))!}{(k+m(n-r+j+1))!}\right)$$

$$= \frac{(mn+k)!}{(mn)!}\ \sum_{j=0}^{r-1}(-1)^j\frac{\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}}{\binom{k+m(n-r+j+1)}{k}}$$

$$= \frac{(mn+k)!}{(mn)!}\ \sum_{j=1}^{r}(-1)^{j-1}\frac{\binom{n-r+j-1}{n-r}\binom{n}{n-r+j}}{\binom{m(n-r+j)+k}{k}}.$$

Now suppose $r = 1, q = 2, k = 1$. Then

$$E\left(W^{[1]}_{1:(2,\ldots,2)}\right) = \frac{1(mn+1)!}{(mn)!}\left(\int_0^1 p^{1-1}\ P((Y_1,\ldots,Y_n)\in\mathbb{A}_{Q:1})\mathrm{d}p\right)$$

$$= (mn+1)\left(\int_0^1 \left(\binom{m}{0}p^0(1-p)^{m-0}+\binom{m}{1}p^1(1-p)^{m-1}\right)^n\mathrm{d}p\right)$$

$$= (mn+1)\left(\int_0^1 (1-p)^{n(m-1)}\ (1-p+mp)^n\mathrm{d}p\right)$$

$$= (mn+1)\left(\int\limits_0^1 (1-p)^{n(m-1)}\ (1+p(m-1))^n \mathrm{d}p\right)$$

$$= (mn+1)\sum_{j=0}^{n}\binom{n}{j}(m-1)^j\left(\int\limits_0^1 p^j(1-p)^{n(m-1)}\mathrm{d}p\right)$$

$$= (mn+1)\sum_{j=0}^{n}\binom{n}{j}(m-1)^j\tfrac{j!(n(m-1))!}{(n(m-1)+j+1)!}$$

$$= \tfrac{mn+1}{n(m-1)+1}\sum_{j=0}^{n}\frac{\binom{n}{j}}{\binom{n(m-1)+j+1}{j}}(m-1)^j$$

$$= \tfrac{1}{\binom{mn}{n}}\sum_{j=0}^{n}\binom{mn+1}{n-j}(m-1)^j$$

$$= \tfrac{1}{\binom{mn}{n}}\sum_{j=0}^{n}\binom{mn+1}{j}(m-1)^{n-j}$$

Now suppose $r = 1, q = 2$.

$$E\left(W_{1:(2,\ldots,2)}^{[k]}\right) = \ \tfrac{k(M+k)!}{M!}\left(\int\limits_0^1 p^{k-1}\ P((Y_1,\ldots,Y_n) \in \mathbb{A}_{Q:r})\mathrm{d}p\right)$$

$$= \tfrac{k(mn+k)!}{(mn)!}\int\limits_0^1 p^{k-1}\left(\sum_{y=0}^{1}\binom{m}{y}p^y(1-p)^{m-y}\right)^n \mathrm{d}p$$

$$= \tfrac{k(mn+k)!}{(mn)!}\int\limits_0^1 p^{k-1}\big(\tbinom{m}{0}p^0(1-p)^{m-0} + \tbinom{m}{1}p^1(1-p)^{m-1}\big)^n\mathrm{d}p$$

$$= \tfrac{k(mn+k)!}{(mn)!}\int\limits_0^1 p^{k-1}\big((1-p)^m + mp(1-p)^{m-1}\big)^n\mathrm{d}p$$

$$= \frac{k(mn+k)!}{(mn)!} \int_0^1 p^{k-1}(1-p)^{n(m-1)} \ (1+p(m-1))^n \mathrm{d}p$$

$$= \frac{k(mn+k)!}{(mn)!} \sum_{j=0}^{n} \binom{n}{j} (m-1)^j \left(\int_0^1 p^{k+j-1}(1-p)^{n(m-1)} \mathrm{d}p \right)$$

$$= \frac{k(mn+k)!}{(mn)!} \sum_{j=0}^{n} \binom{n}{j} (m-1)^j \left(\frac{(k+j-1)!(n(m-1))!}{(n(m-1)+k+j)!} \right)$$

$$= \frac{k(mn+k)!}{(mn)!} \sum_{j=0}^{n} \binom{n}{j} (m-1)^j \left(\frac{(k+j-1)!(n(m-1))!}{(n(m-1)+k+j)!} \right)$$

$$= \frac{k(mn+k)!}{(mn)!(n(m-1)+1)} \sum_{j=0}^{n} \frac{\binom{n}{j}}{\binom{n(m-1)+k+j}{k+j-1}} (m-1)^j$$

Now suppose $q = 2$.

$$\mathrm{E}\left(W^{[k]}_{r:(2,\ldots,2)}\right) = \frac{k(mn+k)!}{(mn)!} \left(\int_0^1 p^{k-1} \ P((Y_1,\ldots,Y_n) \in \mathbb{A}_{Q:r}) \mathrm{d}p \right)$$

$$= \frac{k(mn+k)!}{(mn)!} \int_0^1 p^{k-1} \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \binom{n}{n-r+j+1} \left(\sum_{y=0}^{1} \binom{m}{y} p^y (1-p)^{m-y} \right)^{n-r+j+1} \mathrm{d}p$$

$$= \frac{k(mn+k)!}{(mn)!} \int_0^1 p^{k-1} \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \binom{n}{n-r+j+1} \left((1-p)^m + mp(1-p)^{m-1}\right)^{n-r+j+1} \mathrm{d}p$$

$$= \frac{k(mn+k)!}{(mn)!} \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \binom{n}{n-r+j+1}$$

$$\left(\int_0^1 p^{k-1} \left((1-p)^m + mp(1-p)^{m-1}\right)^{n-r+j+1} \mathrm{d}p \right)$$

$$= \frac{k(mn+k)!}{(mn)!} \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \binom{n}{n-r+j+1}$$

$$\times\left(\int_0^1 p^{k-1}(1-p)^{(m-1)(n-r+j+1)}(1+p(m-1))^{n-r+j+1}\mathrm{d}p\right)$$

$$= \tfrac{k(mn+k)!}{(mn)!}\sum_{j=0}^{r-1}(-1)^j\tbinom{n-r+j}{n-r}\binom{n}{n-r+j+1}$$
$$\times \sum_{i=0}^{n-r+j+1}\tbinom{n-r+j+1}{i}(m-1)^i\left(\int_0^1 p^{k+i-1}(1-p)^{(m-1)(n-r+j+1)}\mathrm{d}p\right)$$

$$= \tfrac{k(mn+k)!}{(mn)!}\sum_{j=0}^{r-1}(-1)^j\tbinom{n-r+j}{n-r}\binom{n}{n-r+j+1}$$
$$\times \sum_{i=0}^{n-r+j+1}\tbinom{n-r+j+1}{i}(m-1)^i\left(\tfrac{(k+i-1)!((m-1)(n-r+j+1))!}{((m-1)(n-r+j+1)+k+i)!}\right)$$

$$= \tfrac{k(mn+k)!}{(mn)!}\sum_{j=0}^{r-1}\sum_{i=0}^{n-r+j+1}(-1)^j\tbinom{n-r+j}{n-r}\binom{n}{n-r+j+1}\tbinom{n-r+j+1}{i}$$
$$\times\left(\tfrac{(k+i-1)!((m-1)(n-r+j+1))!}{((m-1)(n-r+j+1)+k+i)!}\right)(m-1)^i$$

$$= \tfrac{k(mn+k)!}{(mn)!}\sum_{j=0}^{r-1}\sum_{i=0}^{n-r+j+1}(-1)^j\tbinom{n-r+j}{n-r}\binom{n}{n-r+j+1}\tbinom{n-r+j+1}{i}$$
$$\times\left(\tfrac{1}{(m-1)(n-r+j+1)+1}\right)\left(\frac{1}{\binom{(m-1)(n-r+j+1)+k+i}{k+i-1}}\right)(m-1)^i$$

Now suppose $q = m$.

$$\mathrm{E}\left(W^{[k]}_{r:(m,\ldots,m)}\right) = \tfrac{k(mn+k)!}{(mn)!}\left(\int_0^1 p^{k-1}\ P((Y_1,\ldots,Y_n)\in\mathbb{A}_{Q:r})\mathrm{d}p\right)$$

$$= \tfrac{k(mn+k)!}{(mn)!}\left(\int_0^1 p^{k-1}\ P((Y_1,\ldots,Y_n)\in\mathbb{A}_{Q:r})\mathrm{d}p\right)$$

$$= \frac{k(mn+k)!}{(mn)!}\left(\int_0^1 p^{k-1}\sum_{j=0}^{r-1}(-1)^j\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}\left(\sum_{y=0}^{m-1}\binom{m}{y}p^y(1-p)^{m-y}\right)^{n-r+j+1}\mathrm{d}p\right)$$

$$= \frac{k(mn+k)!}{(mn)!}\left(\int_0^1 p^{k-1}\sum_{j=0}^{r-1}(-1)^j\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}(1-p^m)^{n-r+j+1}\mathrm{d}p\right)$$

$$= \frac{k(mn+k)!}{(mn)!}\left(\int_0^1 p^{mi+k-1}\sum_{j=0}^{r-1}\sum_{i=0}^{n-r+j+1}(-1)^{j+i}\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}\binom{n-r+j+1}{i}\mathrm{d}p\right)$$

$$= \frac{k(mn+k)!}{(mn)!}\sum_{j=0}^{r-1}\sum_{i=0}^{n-r+j+1}(-1)^{j+i}\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}\binom{n-r+j+1}{i}\left(\int_0^1 p^{mi+k-1}\mathrm{d}p\right)$$

$$= \frac{k(mn+k)!}{(mn)!}\sum_{j=0}^{r-1}\sum_{i=0}^{n-r+j+1}(-1)^{j+i}\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}\binom{n-r+j+1}{i}\frac{1}{mi+k}$$

$$= \frac{k(mn+k)!}{(mn)!}\sum_{j=0}^{r-1}(-1)^j\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}\left(\sum_{i=0}^{n-r+j+1}(-1)^i\binom{n-r+j+1}{i}\frac{1}{mi+k}\right)$$

$$= \frac{k(mn+k)!}{(mn)!}\sum_{j=0}^{r-1}(-1)^j\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}\left(\frac{m^{n-r+j+1}(n-r+j+1)!}{\prod_{i=0}^{n-r+j+1}(mi+k)}\right)$$

$$= \frac{(mn+k)!}{(mn)!}\sum_{j=0}^{r-1}(-1)^j\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}\left(\frac{m^{n-r+j+1}(n-r+j+1)!}{\prod_{i=1}^{n-r+j+1}(mi+k)}\right)$$

$$= \frac{(mn+k)!}{(mn)!}\sum_{j=0}^{r-1}(-1)^j\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}\left(\frac{(n-r+j+1)!}{\prod_{i=1}^{n-r+j+1}(mi+k)}\right)m^{n-r+j+1}$$

$$= \frac{(mn+k)!}{(mn)!}\sum_{j=0}^{r-1}(-1)^j\binom{n-r+j}{n-r}\binom{n}{n-r+j+1}\left(\prod_{i=1}^{n-r+j+1}\frac{mi}{mi+k}\right)$$

Problem 5.

Suppose we draw t balls without replacement from an urn containing m copies of each of n different colored balls. Let V_r equal the number of colors which are selected r times. Find $E(V_r)$.

Answer

$$n \cdot \frac{\binom{m}{r}\binom{mn-m}{t-r}}{\binom{mn}{t}}$$

Proof

$$\begin{aligned}
E(V_r) &= E\left(\sum_{j=0}^{n} \mathrm{I}_{\{r\}}(X_j)\right) = \sum_{j=0}^{n} E\big(\mathrm{I}_{\{r\}}(X_j)\big) \\
&= \sum_{j=0}^{n} P(X_j = r) \\
&= \sum_{j=0}^{n} \frac{\binom{m}{r}\binom{mn-m}{t-r}}{\binom{mn}{t}} \\
&= n \cdot \frac{\binom{m}{r}\binom{mn-m}{t-r}}{\binom{mn}{t}} \qquad \square
\end{aligned}$$

Problem 5.

Show that

$$\binom{mn}{t} = \sum_{i=0}^{n} \sum_{j=0}^{n-i} (-1)^j \binom{mn - m(i+j)}{t - r(i+j)} \binom{n}{i} \binom{n-i}{j} \binom{m}{r}^{i+j}$$

and

$$n\binom{m}{r}\binom{mn-m}{t-r} = \sum_{i=0}^{n} \sum_{j=0}^{n-i} (-1)^j \binom{mn - m(i+j)}{t - r(i+j)} \binom{n}{i} \binom{n-i}{j} \binom{m}{r}^{i+j} (i)$$

Proof

By the law of total probability

$$1=\sum_{i=0}^{n} P(V_r=i)=\sum_{i=0}^{n}\ \sum_{j=0}^{n-i}\ (-1)^j \frac{\binom{n}{i}}{\binom{mn}{t}}\binom{mn-m(i+j)}{t-r(i+j)}\binom{n-i}{j}\binom{m}{r}^{i+j}$$

from which the first identity follows immediately.

To establish the second identity, we note from Problems 1 and 2 taken together that,

$$\begin{aligned}\frac{n}{\binom{mn}{t}}\binom{m}{r}\binom{mn-m}{t-r} &= E(V_r)=\sum_{i=0}^{n} iP(V_r=i)\\ &=\sum_{i=0}^{n} i\frac{\binom{n}{i}\binom{m}{r}^i}{\binom{mn}{t}}\left(\sum_{j=0}^{n-i}\ (-1)^j\binom{mn-m(i+j)}{t-r(i+j)}\binom{n-i}{j}\binom{m}{r}^j\right)\\ &=\sum_{i=0}^{n}\ \sum_{j=0}^{n-i} (-1)^j\, i\frac{\binom{n}{i}}{\binom{mn}{t}}\binom{mn-m(i+j)}{t-r(i+j)}\binom{n-i}{j}\binom{m}{r}^{i+j}\end{aligned}$$

It follows that

$$n\binom{m}{r}\binom{mn-m}{t-r}=\sum_{i=0}^{n}\ \sum_{j=0}^{n-i}\ (-1)^j\binom{mn-m(i+j)}{t-r(i+j)}\binom{n}{i}\binom{n-i}{j}\binom{m}{r}^{i+j}\ (i)$$

□

<u>Problem 6a.</u>

Suppose we draw n balls without replacement from an urn containing r copies of each of m different colored balls. Let Z equal the number of different colors that are selected in this sample.

$$P(Z=k)=\frac{m_{(k)}}{(rm)_{(n)}}C(n,k,r)$$

where

$$C(n,k,r)=\frac{1}{k!}\sum_{j=0}^{k}(-1)^{k-j}\binom{k}{j}(rj)_{(n)}$$

$$\mathrm{E}\Big((m-Z)_{(s)}\Big) = m_{(s)}\frac{(r(m-s))_{(n)}}{(rm)_{(n)}}$$

References:

Walton

Korwar

Charalambides, “On a Restricted Occupancy Model and its Applications”, *Biom. Journal*, Vol. 23, no. 6, 1981, pages 601-610.

Problem 6b.

Suppose we draw n balls without replacement from an urn containing r copies of each of m different solid colored balls and s identical striped balls. Let U equal the number of different solid colors that are selected in this sample and let V be the number of different striped balls that are selected in this sample.

$$P(U=k,\, V=n-j) = \binom{n}{j}\frac{s_{(n-j)}(rm)_{(j)}}{(rm+s)_{(n)}}G(n,k,r,s)$$

where

$G(n,k,r,s) =$

$$\mathrm{E}\Big((m-U)_{(\nu)}(V)_{(\tau)}\Big) = \frac{m_{(\nu)}(rm-r\nu+s)_{(n)}}{(rm{+}s)_{(n)}}\ \frac{n_{(\tau)}(rm-r\nu)_{(\tau)}}{(rm-r\nu+s)_{(\tau)}}$$

$$\mathrm{E}\Big((m-U)_{(\nu)}\Big) = \frac{m_{(\nu)}(rm-r\nu+s)_{(n)}}{(rm{+}s)_{(n)}}$$

References:

Charalambides, "On a Restricted Occupancy Model and its Applications", *Biom. Journal*, Vol. 23, no. 6, 1981, pages 601-610.

Problem 7.

We have $n + r$ distinguishable urns, each with s distinguishable cells. A cell cannot hold more than ball. Identical balls are randomly distributed (all empty cells equally likely at each turn) until k urns, among the n specified urns, are occupied by at least one ball. Let M equal the number of turns required.

$$P(M = m) = \frac{s\,n_{(k)}}{(sn + sr)_{(m)}} C(m - 1, k - 1, s, rs)$$

for $m = k, k + 1, \ldots$

where

$$C(m - 1, k - 1, s, rs) = \frac{1}{(k - 1)!} \sum_{j=0}^{k-1} (-1)^{k-j-1} \binom{k - 1}{j} (sj + sr)_{(m-1)}$$

is the non-central C-number (Charalambides, Koutras, "On the Differences of the Generalized Factorials at an Arbitrary Point and Their Combinatorial Applications", *Discrete Mathematics*, Vol. 47, 1983, pages 183 - 201.)

Charalambides, Ch. A., "A Unified Derivation of Occupancy and Sequential Occupancy Distributions", Advances in Combinatorial Methods and Applications to Probability and Statistics, N. Balakrishnan (editor), 1997, pages 259 - 273

$$\mathrm{E}\left(M^{(j)}\right) = \binom{n - 1}{k - 1} \sum_{i=0}^{k-1} (-1)^{k-i-1} \binom{k - 1}{i} \frac{j!sn(sn + sr + j)_{(j)}}{(sn - si + j)_{(j+1)}}$$

Problem 8.

An urn consists of $m = 2$ balls of each of s different colors. Balls are drawn without replacement until both balls of some color have been drawn out. Let N equal the number of draws required.

$$\mathrm{E}(N) = \frac{(2s)!!}{(2s-1)!!}$$

$$\mathrm{E}(N(N+1)) = 2(2s+1)$$

Reference:

Blom, Gunnar and Holst, Lars, “Embedding Procedures for Discrete Problems in Probability”, *Mathematical Scientist*, 16, 29-40, 1991.

Problem 9.

Suppose we draw t balls without replacement from an urn containing m copies of each of n different colored balls. Let D_r equal the number of colors which are selected exactly r times. The joint descending factorial moment $\mathrm{E}\Big((D_0)_{(r_0)}\cdots(D_t)_{(r_t)}\Big)$ is given by

$$\mathrm{E}\Big((D_0)_{(r_0)}\cdots(D_t)_{(r_t)}\Big) = \frac{n_{(R)}\binom{m(n-R)}{t-S}}{\binom{mn}{t}}\prod_{j=0}^{t}\binom{m}{j}^{r_j}$$

where $R = r_0 + r_1 + \ldots + r_t$ and $S = 0r_0 + 1r_1 + \ldots + tr_t$. The special case $\mathrm{E}\Big((D_0)_{(r_0)}\Big)$ (*i.e.* $r_1 = \ldots = r_t = 0$) simplifies to

$$\mathrm{E}\Big((D_0)_{(r_0)}\Big) = n_{(r_0)}\frac{(m(n-r_0))_{(t)}}{(mn)_{(t)}}$$

in agreement with C. Charalambides, On a Restricted Occupancy Model and its Applications, *Biom. Journal* **23** (1981), no. 6, 601-610.

Proof

$$\mathrm{E}\Big((D_0)_{(r_0)}\cdots(D_t)_{(r_t)}\Big) = \frac{1}{\binom{mn}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\Big(e^{\theta(1+\lambda)^m}\mathrm{E}\Big((Z_0)_{(r_0)}\cdots(Z_t)_{(r_t)}\Big)\Big)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{1}{\binom{mn}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\left(e^{\theta(1+\lambda)^m}\prod_{j=0}^{t}\mathrm{E}\Big((Z_j)_{(r_j)}\Big)\right)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{1}{\binom{mn}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\left(e^{\theta(1+\lambda)^m}\prod_{j=0}^{t}\left(\binom{m}{j}\theta\lambda^j\right)^{r_j}\right)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{\prod_{j=0}^{t}\binom{m}{j}^{r_j}}{\binom{mn}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\left(e^{\theta(1+\lambda)^m}\theta^R\lambda^S\right)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

where $R = r_0 + r_1 + \ldots + r_t$ and $S = 0r_0 + 1r_1 + \ldots + tr_t$

$$= \frac{\prod_{j=0}^{t}\binom{m}{j}^{r_j}}{\binom{mn}{t}t!}\frac{n!}{(n-R)!}\frac{d^t}{d\lambda^t}\Big((1+\lambda)^{m(n-R)}\lambda^S\Big)\Big|_{\lambda=0}$$

$$= \frac{\prod_{j=0}^{t}\binom{m}{j}^{r_j}}{\binom{mn}{t}t!}\frac{n!}{(n-R)!}\binom{m(n-R)}{t-S}t!$$

$$= \frac{n_{(R)}\binom{m(n-R)}{t-S}}{\binom{mn}{t}}\prod_{j=0}^{t}\binom{m}{j}^{r_j}.$$

Method 2

Blom, Gunnar and Holst, Lars, "*Embedding Procedures for Discrete Problems in Probability*", **Mathematical Scientist**, 16, 29-40, 1991.

Let $\Big((Y_{1,1},\ldots,Y_{1,m_1}),\ldots,(Y_{i,1},\ldots,Y_{i,m_i}),\ldots,(Y_{n,1},\ldots,Y_{n,m_n})\Big)$ be a random sample of size $M = m_1 + \ldots + m_n$ from the continuous Uniform$(0,1)$ distribution.

Let $Y_{(1)} < Y_{(2)} < \ldots < Y_{(M)}$ be the ordered values of $Y_{i,j}$.

We will say that $Y_{(j)}$ is from the i^{th} group provided $Y_{(j)} \in (Y_{i,1},\ldots,Y_{i,m_i})$.

For $1 \le t \le M$, let $N_{i,t}$ equal the number of values among $Y_{(1)} < \ldots < Y_{(t)}$ which are from the i^{th} group.

Now consider drawing objects one by one at random and without replacement from an urn initially containing m_i objects of color i, $i = 1,\ldots,n$.

For $1 \le t \le M$, let $C_{i,t}$ equal the number of objects of color i drawn in the first t draws from this urn.

"Clearly" for all $1 \le t \le M$,

$$\Big(N_{1,t},\ldots,N_{n,t}\Big) \overset{d}{=} \Big(C_{1,t},\ldots,C_{n,t}\Big).$$

When q_i objects of color i have been drawn $\Big($or equivalently when q_i values from the i^{th} group of uniform variates $(Y_{i,1},\ldots,Y_{i,m_i})$ have appeared in the list $Y_{(1)} < Y_{(2)} < \ldots < Y_{(M)}\Big)$ we will say this color (or group) has reached its *quota*.

Let $\mathbb{C}(k;q_1,\ldots,q_n) = \mathbb{C}_k$ represent the draw (index) when exactly k non-specified colors (groups) reach their quota.

Our goal is to find a formula for $E\Big(\mathbb{C}_k\Big)$.

Define $Y_{(0)} = 0$. We note that

$$Y_{(\mathbb{C}_k)} = \sum_{j=1}^{\mathbb{C}_k} D_j$$

where $D_j = Y_{(j)} - Y_{(j-1)}$. Furthermore "it is well know" that $D_1,\ldots,D_n$ are "exchangeable" random variables following a Beta distribution with parameters 1 and $M = m_1 + \ldots + m_n$.

That is the density function $f(y)$ of D_j is

$$f(y) = M(1 - y)^{M-1}.$$

Now let $\mathrm{Y}_{(i,1)} < \mathrm{Y}_{(i,2)} < \ \ldots < \ \mathrm{Y}_{(i,m_i)}$ be the ordered values of the *iid* Uniform(0,1) variates $(\mathrm{Y}_{i,1},\ldots,\mathrm{Y}_{i,m_i})$.

It follows that

$\mathrm{Y}_{(\mathbb{C}_k)} \ = \ k^{th}$ largest value in the set of ***independent*** variates $\left\{\mathrm{Y}_{(1,q_1)},\, \mathrm{Y}_{(2,q_2)},\, \ldots,\, \mathrm{Y}_{(n,q_n)}\right\}$.

However “it is well know" that the j^{th} largest value from a set of *n iid* Uniform(0,1) variates follows a Beta distribution with parameters q_j and $m_j - q_j + 1$.

That is the density function f(y) of $\mathrm{Y}_{(j,q_j)}$ is

$$\mathrm{f}(y) \ = \ \frac{m_j!}{(q_j-1)!\,(m_j-q_j)!}\, y^{q_j-1}(1-\ y)^{m_j-q_j}.$$

Finally “we can show" that $\mathbb{C}_k$ is a “stopping time". Hence

$$\mathrm{E}\left(\mathrm{Y}_{(\mathbb{C}_k)}\right)$$

$$= \mathrm{E}\left(\sum_{j=1}^{\mathbb{C}_k} \mathrm{D}_j\right)$$

$$= \mathrm{E}\left(\mathbb{C}_k\right) \ \cdot \ \mathrm{E}\left(\mathrm{D}_1\right)$$

$$= \mathrm{E}\left(\mathbb{C}_k\right) \ \cdot \ \frac{1}{\mathrm{M}+1}$$

Therefore,

$$\mathrm{E}\left(\mathbb{C}_k\right) \ = \ (\mathrm{M}+\ 1) \ \cdot \ \mathrm{E}\left(\mathrm{Y}_{(\mathbb{C}_k)}\right)$$

where

$\mathrm{Y}_{(\mathbb{C}_k)}$ is the k^{th} largest value in the set of ***independent*** Beta variates $\left\{\mathrm{Y}_{(1,q_1)},\, \mathrm{Y}_{(2,q_2)},\, \ldots,\, \mathrm{Y}_{(n,q_n)}\right\}$

where $\mathrm{Y}_{(i,q_i)} \ \sim$ Beta distribution $(q_j,\, m_j - q_j + \ 1)$.

In our particular problem

$$m_1 = \ldots = m_n = m \quad \text{and} \quad q_1 = \ldots = q_n = 1.$$

For the purposes of Theorem 3 which follows we will suppose that $Y_1,\ldots,Y_n$ are independent **binomial** random variables parameterized such that for $j = 1,2,\ldots,n$

$$P(Y_j = y) = \binom{m_j}{y} p^y (1-p)^{m_j - y} \qquad y = 0, 1, \ldots, m_j.$$

Suppose an urn contains m_1 objects of Type $1,\ldots,m_n$ objects of Type n and that objects are drawn from this urn without replacement. Let $M = m_1 + \ldots + m_n$. When q_i objects of Type i have been drawn we will say Type i has reached its ***quota***.

Let $W_{(r:q_1,\ldots,q_n)}(m_1,\ldots,m_n;n) \equiv W_{r:Q}$ represent the waiting time until exactly r different types have reached their quota.

Let $E\left(W_{r:Q}^{[k]}\right)$ represent the k^{th} ascending moment of $W_{r:Q}$. That is,

$$E\left(W_{r:Q}^{[k]}\right) = E((W_{r:Q}+0)(W_{r:Q}+1)\cdots(W_{r:Q}+k-1))$$

Theorem 3.

$$E\left(W_{r:Q}^{[k]}\right) = \frac{k(M+k)!}{M!}\left(\int_0^1 p^{k-1}\, P((Y_1,\ldots,Y_n) \in \mathbb{A}_{Q:r})\mathrm{d}p\right)$$

where

(1) $\mathbb{A}_{Q:r}$ is the event that at least $n-r+1$ of the (independent) events $\mathcal{A}_1,\ldots,\mathcal{A}_n$ occur

(2) $\mathcal{A}_j$ is the event that $Y_j < q_j$.

$P((Y_1,\ldots,Y_n) \in \mathbb{A}_{Q:r}) = P($at least $n-r+1$ of the events $\mathcal{A}_1,\ldots,\mathcal{A}_n$ occur$)$

$= P($at least $n-r+1$ of the $Y_j < q_j$ $)$

$= \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \mathbb{S}_{n-r+j+1}$

$$= \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \binom{n}{n-r+j+1} (P(Y_j < q))^{n-r+j+1}$$

$$= \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \binom{n}{n-r+j+1} \left(\sum_{y=0}^{q-1} \binom{m}{y} p^y (1-p)^{m-y} \right)^{n-r+j+1}$$

$$= \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \binom{n}{n-r+j+1} \left(\sum_{y=0}^{q-1} \binom{m}{y} p^y (1-p)^{m-y} \right)^{n-r+j+1}$$

$$\int_0^1 p^{k-1} \; P((Y_1,\ldots,Y_n) \in \mathbb{A}_{Q:r}) \mathrm{d}p$$

$$= \int_0^1 p^{k-1} \sum_{j=0}^{r-1} (-1)^j \binom{n-r+j}{n-r} \binom{n}{n-r+j+1} \left(\sum_{y=0}^{q-1} \binom{m}{y} p^y (1-p)^{m-y} \right)^{n-r+j+1} \mathrm{d}p$$

Suppose $r = 1$.

$$\int_0^1 p^{k-1} \left(\sum_{y=0}^{q-1} \binom{m}{y} p^y (1-p)^{m-y} \right)^{n} \mathrm{d}p$$

If $q = 1$

$$\int_0^1 p^{k-1} \left(\binom{m}{0} p^0 (1-p)^{m-0} \right)^n \mathrm{d}p$$

$$= \int_0^1 p^{k-1} (1-p)^{mn} \mathrm{d}p$$

$$= \sum_{j=0}^{mn} (-1)^j \binom{mn}{j} \int_0^1 p^{k-1+j} \mathrm{d}p$$

$$= \sum_{j=0}^{mn} (-1)^j \binom{mn}{j} \left(\frac{1}{k+j} \right)$$

$= \frac{k}{\binom{mn+k}{mn}}$ using Gould's identity (1.41)

If $q = 2$

$$\int_0^1 p^{k-1}\left(\sum_{y=0}^{1}\binom{m}{y}p^y(1-p)^{m-y}\right)^n \mathrm{d}p$$

$$= \int_0^1 p^{k-1}\left(\binom{m}{0}p^0(1-p)^m + \binom{m}{1}p^1(1-p)^{m-1}\right)^n \mathrm{d}p$$

$$= \int_0^1 p^{k-1}\left((1-p)^m + mp(1-p)^{m-1}\right)^n \mathrm{d}p$$

$$= \int_0^1 p^{k-1}(1-p)^{m-1}(1+(m-1)p)^n \mathrm{d}p$$

$$= \sum_{j=0}^{n}\binom{n}{j}(m-1)^j \int_0^1 p^{k+j-1}(1-p)^{m-1}\mathrm{d}p$$

$$= \sum_{j=0}^{n}\binom{n}{j}(m-1)^j \frac{(k+j-1)!(m-1)!}{(k+j+m-1)!}$$

$$= \frac{1}{m}\sum_{j=0}^{n} (m-1)^j \frac{\binom{n}{j}}{\binom{m+k+j-1}{m}}$$

For the special case $k = 1$, this simplifies to

$$\frac{1}{m}\sum_{j=0}^{n} (m-1)^j \frac{\binom{n}{j}}{\binom{m+j}{m}}$$

$$= \frac{1}{m\binom{m+n}{m}}\sum_{j=0}^{n} (m-1)^j \binom{m+n}{m+j}$$

$$=$$

However, it is well known that for nonnegative integers a and b,

$$\int_0^1 p^a(1-p)^b \mathrm{d}p = \frac{a!b!}{(a+b+1)!}$$

$$P(Y_j < q) = \sum_{k=0}^{q-1} \binom{m}{y} p^y (1-p)^{m-y}$$

$$\mathbb{S}_{n-r+j+1} = \begin{cases} \binom{n}{n-r+j+1} (P(Y_j < q))^{n-r+j+1} & 1 \le n-r+j+1 \le n \\ 1 & n-r+j+1 = 0 \end{cases}$$

$$P(\mathbb{H}_{\ge m}) = \sum_{j=0}^{n-m} (-1)^j \binom{m-1+j}{m-1} \mathbb{S}_{m+j}$$

Suppose $A_1, A_2, \ldots, A_n$ are sets within a universal set Ω. Define :

$$\mathbb{H}_{\ge m} = \{x \in \Omega \mid x \text{ is an element of \underline{at least} } m \text{ of the } n \text{ sets } A_1, A_2, \ldots, A_n\}$$

Define

$$\mathbb{S}_k = \begin{cases} \sum_{(j_1, \ldots, j_k) \varepsilon \mathbb{C}_k} P(A_{j_1} \cap \cdots \cap A_{j_k}) & 1 \le k \le n \\ 1 & k = 0 \end{cases}$$

where $\mathbb{C}_k$ is the set of all samples of size k drawn without replacement from $\{1, 2, \ldots, n\}$, where the order of sampling is not considered important. Then,

$$P(\mathbb{H}_{\ge m}) =$$

$\binom{m-1}{m-1}\mathbb{S}_m - \binom{m}{m-1}\mathbb{S}_{m+1} + \ldots + (-1)^{k-m}\binom{k-1}{m-1}\mathbb{S}_k + \ldots + (-1)^{n-m}\binom{n-1}{m-1}\mathbb{S}_n.$

$= \sum_{j=0}^{n-m} (-1)^j \binom{m-1+j}{m-1}\mathbb{S}_{m+j}$

Sampling Without Replacement Model or Grouped Fermi-Dirac Allocation Model

Problem 3.

Define $\mathbb{S}^n$ to be the product space $\{0,1,\ldots,m_1\} \times \cdots \times \{0,1,\ldots,m_n\}$ and let $\mathbb{S}_t^n$ be the set of all vectors $(s_1,\ldots,s_n)$ in $\mathbb{S}^n$ such that $s_1 + \ldots + s_n = t$.

Suppose that $(X_1,\ldots,X_n)$ is a **multivariate hypergeometric** random vector. That is, for all $(s_1,\ldots,s_n) \in \mathbb{S}_t^n$,

$$P(X_1 = s_1,\ldots,X_n = s_n) = \frac{\binom{m_1}{s_1}\cdots\binom{m_n}{s_n}}{\binom{M}{t}}$$

where $M = m_1 + \ldots + m_n$ and $t = s_1 + \ldots + s_n$.

For the purposes of Theorems 1 and 2 which follow we we will suppose that $Y_1,\ldots,Y_n$ are independent **binomial** random variables parameterized such that for $j = 1,2,\ldots,n$

$$P(Y_j = y) = \binom{m_j}{y}\left(\frac{\theta}{\theta+1}\right)^y\left(1 - \frac{\theta}{\theta+1}\right)^{m_j - y} \qquad y = 0, 1, \ldots, m_j.$$

Theorem 1.

$$E_t(g(X_1,\ldots,X_n)) = \frac{1}{\binom{M}{t}t!}\frac{d^t}{d\theta^t}\Big((1+\theta)^M E(g(Y_1,\ldots,Y_n))\Big)\Big|_{\theta=0}$$

where the t in $E_t(\,\cdot\,)$ is used to denote that $X_1 + \ldots + X_n = t$.

Theorem 2.

Let $\mathcal{A} \subset \mathbb{S}^n$ and define $\mathcal{A}_t = \mathcal{A} \cap \mathbb{S}_t^n$. Then for $t \geq 0$,

$$P((X_1,\ldots,X_n) \in \mathcal{A}_t) = \frac{1}{\binom{M}{t} t!} \frac{d^t}{d\theta^t}\left((1+\theta)^M P((Y_1,\ldots,Y_n) \in \mathcal{A})\right)\Big|_{\theta=0}$$

For the purposes of Theorem 3 which follows we will suppose that $Y_1,\ldots,Y_n$ are independent **binomial** random variables parameterized such that for $j = 1,2,\ldots,n$

$$P(Y_j = y) = \binom{m_j}{y} p^y (1-p)^{m_j - y} \qquad y = 0, 1, \ldots, m_j.$$

Suppose an urn contains m_1 objects of Type $1,\ldots,m_n$ objects of Type n and that objects are drawn from this urn without replacement. Let $M = m_1 + \ldots + m_n$. When q_i objects of Type i have been drawn we will say Type i has reached its ***quota***.

Let $W_{(r:q_1,\ldots,q_n)}(m_1,\ldots,m_n;n) \equiv W_{r:Q}$ represent the waiting time until exactly r different types have reached their quota.

Let $E\left(W_{r:Q}^{[k]}\right)$ represent the k^{th} ascending moment of $W_{r:Q}$. That is,

$$E\left(W_{r:Q}^{[k]}\right) = E((W_{r:Q}+0)(W_{r:Q}+1)\cdots(W_{r:Q}+k-1))$$

<u>**Theorem 3.**</u>

$$E\left(W_{r:Q}^{[k]}\right) = \frac{k(M+k)!}{M!}\left(\int_0^1 p^{k-1}\, P((Y_1,\ldots,Y_n) \in \mathbb{A}_{Q:r})\mathrm{d}p\right)$$

and

$$E\left(W_{r:Q}^{[k]} - W_{r-1:Q}^{[k]}\right) = \frac{k(M+k)!}{M!}\left(\int_0^1 p^{k-1}\, P((Y_1,\ldots,Y_n) \in \mathbb{B}_{Q:r})\mathrm{d}p\right)$$

where

(1) $\mathbb{A}_{Q:r}$ is the event that at least $n-r+1$ of the (independent) events $\mathcal{A}_1,\ldots,\mathcal{A}_n$ occur

$\mathbb{B}_{Q:r}$ is the event that exactly $n-r+1$ of the (independent) events $\mathcal{A}_1,\ldots,\mathcal{A}_n$ occur

(2) $\mathcal{A}_j$ is the event that $Y_j < q_j$.

Problem 1.

(a) Suppose we draw t balls without replacement from an urn containing m copies of each of n different colored balls. The probability that exactly v colors are selected exactly r times equals

$$\frac{1}{\binom{mn}{t}} \sum_{j=v}^{n} (-1)^{j-v} \binom{mn-mj}{t-rj} \binom{j}{v} \binom{n}{j} \binom{m}{r}^j$$

provided $t \in \{rv, \ldots, rn\}$.

References

The special case of $v = 0$ and $r = 0$ is equation 6.1 of Charalambides, "A New Kind of Numbers Appearing in the n-Fold Convolution of Truncated Binomial and Negative Binomial Distributions", *SIAM Journal of Applied Mathematics*, Vol 33, No. 2, September 1977, pages 279-288. Charalambides's expresses his solution in the form

$$\frac{n!}{(nm)_{(t)}} \mathrm{C}(t, n, m)$$

where $\mathrm{C}(t, n, m)$ are the C-numbers defined by

$$\mathrm{C}(t, n, m) = \frac{1}{n!} \sum_{j=0}^{n} (-1)^{n-j} \binom{n}{j} (mj)_{(t)}$$

(b) The probability that at least v colors are selected exactly r times equals

$$\frac{1}{\binom{mn}{t}} \sum_{j=v}^{n} (-1)^{j-v} \binom{mn-mj}{t-rj} \binom{j-1}{v-1} \binom{n}{j} \binom{m}{r}^j$$

provided $t \in \{rv, \ldots, rn\}$.

(c) The expected number of colors that are drawn exactly r times equals

$$\frac{n\binom{m}{r}\binom{mn-m}{t-r}}{\binom{mn}{t}}$$

<u>Problem 2.</u>

Suppose we continue to draw balls without replacement from an urn containing m copies of each of n different colored balls.

(a) The probability that w draws will be required to secure at least one ball of every color equals

$$\frac{n!m}{(nm)_w}\mathrm{C}(w-1,n-1,m)$$

<u>References</u>

This result is equation 6.2 of Charalambides, "A New Kind of Numbers Appearing in the n-Fold Convolution of Truncated Binomial and Negative Binomial Distributions", *SIAM Journal of Applied Mathematics*, Vol 33, No. 2, September 1977, pages 279-288.

(b) The probability that w draws will be required to secure at least two balls of every color equals

$$\frac{n!m_{(2)}(w-1)^2}{(mn)_{(w)}}\sum_{j=1}^{n-1}(-1)^{j-1}m^j\binom{w-2}{j}C(w-j-2,n-j-1,m)$$

<u>Note</u>

The result in 2(b) can be derived from 2(a) by an inclusion-exclusion argument and the waiting time to secure at least <u>three</u> balls of every color could be derived from 2(b) by the same inclusion-exclusion argument. Unfortunately continuing in this way does not seem to lead to a succinct formula for the waiting time to secure at least k balls of every color.

The probability that w draws will be required to secure k balls of every color can be expressed in terms of the *Generalized C-Numbers* (Equation 3.13, Charalambides, "The Generalized Stirling and C numbers", *Sankhyā*, *Series A*, Vol. 36, Pt. 4, 1974, pp. 419-436), but there is no succinct formula for the generalized C-numbers, even though many properties and applications of these numbers are well known.

(c) The k^{th} ascending factorial moment of the number of draws required to secure r of the n different colored balls equals

$$\frac{(mn)!}{(mn+k)!}\sum_{j=1}^{r}(-1)^{j-1}\frac{\binom{n-r+j-1}{n-r}\binom{n}{n-r+j}}{\binom{m(n-r+j)+k}{k}}$$

The special case $k=1$ and $r=n$ simplifies to

$$1+nm\left(1-\prod_{j=1}^{n-1}\frac{mj}{mj+1}\right)$$

We can compare this result with its well known formula with replacement analog

$$n\left(\frac{1}{1}+\frac{1}{2}+\ldots+\frac{1}{n}\right)$$

for the expected number of draws required to secure all n different colored balls when sampling with replacement from an urn with n different colored balls.

Problem 3.

(a) Suppose we draw n balls without replacement from an urn containing r copies of each of m different solid colored balls and s identical striped balls. Let U equal the number of different solid colors that are selected in this sample and let V be the number of different striped balls that are selected in this sample.

$$P(U=k,\, V=n-j)=\binom{n}{j}\frac{s_{(n-j)}(rm)_{(j)}}{(rm+s)_{(n)}}G(n,k,r,s)$$

where $G(n,k,r,s)$ are the Gould-Hopper numbers.

(b)

$$\mathrm{E}\Big((m-U)_{(\nu)}(V)_{(\tau)}\Big) = \frac{m_{(\nu)}(rm-r\nu+s)_{(n)}}{(rm+s)_{(n)}}\ \frac{n_{(\tau)}(rm-r\nu)_{(\tau)}}{(rm-r\nu+s)_{(\tau)}}$$

(c)

$$\mathrm{E}\Big((m-U)_{(\nu)}\Big) = \frac{m_{(\nu)}(rm-r\nu+s)_{(n)}}{(rm+s)_{(n)}}$$

References

Charalambides, "On a Restricted Occupancy Model and its Applications", *Biom. Journal,* Vol. 23, no. 6, 1981, pages 601-610.

Problem 4.

We have $n+r$ distinguishable urns, each with s distinguishable cells. A cell cannot hold more than ball. Identical balls are randomly distributed (all empty cells equally likely at each turn) until k urns, among the n specified urns, are occupied by at least one ball. Let M equal the number of turns required.

(a)

$$P(M=m) = \frac{s\,n_{(k)}}{(sn+sr)_{(m)}} C(m-1,k-1,s,rs)$$

for $m = k, k+1, \ldots$

where

$$C(m-1,k-1,s,rs) = \frac{1}{(k-1)!}\sum_{j=0}^{k-1}(-1)^{k-j-1}\binom{k-1}{j}(sj+sr)_{(m-1)}$$

is the non-central C-number (Charalambides, Koutras, "On the Differences of the Generalized Factorials at an Arbitrary Point and Their Combinatorial Applications", *Discrete Mathematics,* Vol. 47, 1983, pages 183 - 201.)

References

Charalambides, Ch. A., "A Unified Derivation of Occupancy and Sequential Occupancy Distributions", Advances in Combinatorial Methods and Applications to Probability and Statistics, N. Balakrishnan (editor), 1997, pages 259 - 273

(b)

$$\mathrm{E}\left(M^{(j)}\right) = \binom{n-1}{k-1} \sum_{i=0}^{k-1} (-1)^{k-i-1} \binom{k-1}{i} \frac{j! sn(sn+sr+j)_{(j)}}{(sn-si+j)_{(j+1)}}$$

<u>References</u>

Charalambides, Ch. A., "<u>A Unified Derivation of Occupancy and Sequential Occupancy Distributions</u>", Advances in Combinatorial Methods and Applications to Probability and Statistics, N. Balakrishnan (editor), 1997, pages 259 - 273

<u>Problem 5.</u>

The k^{th} ascending factorial moment of the number of draws required to draw out all m copies of any color of ball when sampling without replacement from an urn containing m copies of each of n different colored balls equals

$$\frac{(mn+k)!}{(mn)!} \prod_{j=1}^{n} \left(\frac{mj}{mj+k} \right)$$

<u>References</u>

The two special cases $m = 2,\ k = 1$ and $m = 2,\ k = 2$ are given in Blom, Gunnar and Holst, Lars, "Embedding Procedures for Discrete Problems in Probability", *Mathematical Scientist*, 16, 29-40, 1991.

Example 7

Suppose we draw n balls without replacement from an urn containing r copies of each of m different solid colored balls and s identical striped balls. Let U equal the number of different solid <u>colors</u> that are selected in this sample and let V be the number of striped <u>balls</u> that are selected in this sample.

(a) $$P(U = k,\, V = n - j) = \binom{n}{j} \frac{s_{(n-j)}(rm)_{(j)}}{(rm+s)_{(n)}} G(n,k,r,s)$$

where $G(n,k,r,s)$ are the Gould-Hopper numbers.

$$G(n,k,r,s) = \frac{n!}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \binom{rj+s}{n}$$

Proof

$$P(U = k,\, V = n - j) = P(C_0 = m - k, 0D_0 + 1D_1 + \ldots + tD_t = n - j)$$

Suppose an urn containins m copies of each of n different colored solid balls and s copies of each of $r = 1$ different colored striped balls.

The number of ways to select t balls without replacement from this urn and get a sample with k of the n solid colors and a total of $t - j$ striped balls equals

$$\binom{s}{t-j}\binom{n}{k}\frac{k!}{j!}C(j,k,m).$$

where the C-numbers where defined earlier by

$$\frac{k!}{j!}C(j,k,m) = \sum_{i=0}^{k}(-1)^{k-i}\binom{k}{i}\binom{im}{j}.$$

Charalambides [] considers this problem. There is a misprint where he uses $C(j,k,s)$ instead of $C(j,k,m)$.

Proof

In the notation of Theorem 6 with $r = 1$, the problem asks us to find

$\binom{mn+s}{t}P(C_0 = n-k, Y_1 = t-j)$. By Theorem 6

$$P(C_0 = n-k, Y_1 = t-j)$$

$$= \frac{1}{\binom{mn+s}{t}t!}\ \frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\big((1+\lambda)^s e^{\theta(1+\lambda)^m}P(W_0 = n-k, Z_1 = t-j)\big)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

where $W_0 \sim$ Poisson(θ) and $Z_1 \sim$ Binomial$\big(s, \frac{\lambda}{1+\lambda}\big)$

$$= \frac{1}{\binom{mn+s}{t}t!}\ \frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\left((1+\lambda)^s e^{\theta(1+\lambda)^m}\frac{e^{-\theta}\theta^{n-k}}{(n-k)!}\binom{s}{t-j}\left(\frac{\lambda}{1+\lambda}\right)^{t-j}\left(\frac{1}{1+\lambda}\right)^{s-(t-j)}\right)\Bigg|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{\binom{s}{t-j}}{\binom{mn+s}{t}t!(n-k)!}\ \frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\left(e^{\theta((1+\lambda)^m-1)}\theta^{n-k}\lambda^{t-j}\right)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{\binom{s}{t-j}}{\binom{mn+s}{t} t!(n-k)!} \frac{d^t}{d\lambda^t}\left(((1+\lambda)^m - 1)^{n-(n-k)} \frac{n!}{(n-(n-k))!} \lambda^{t-j} \right)\Big|_{\lambda=0}$$

$$= \frac{\binom{s}{t-j}}{\binom{mn+s}{t} t!(n-k)!} \frac{n!}{k!} \frac{d^t}{d\lambda^t}\left(((1+\lambda)^m - 1)^{k} \lambda^{t-j} \right)\Big|_{\lambda=0}$$

$$= \frac{\binom{s}{t-j}\binom{n}{k}}{\binom{mn+s}{t} t!} \frac{d^t}{d\lambda^t}\left(((1+\lambda)^m - 1)^{k} \lambda^{t-j} \right)\Big|_{\lambda=0}$$

$$= \frac{\binom{s}{t-j}\binom{n}{k}}{\binom{mn+s}{t} t!} \frac{d^t}{d\lambda^t}\left(\sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} (1+\lambda)^{im} \lambda^{t-j} \right)\Big|_{\lambda=0}$$

$$= \frac{\binom{s}{t-j}\binom{n}{k}}{\binom{mn+s}{t} t!} \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} \binom{im}{t-(t-j)} t!$$

$$= \frac{\binom{s}{t-j}\binom{n}{k}}{\binom{mn+s}{t}} \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} \binom{im}{j}$$

Therefore,

$$\binom{mn+s}{t} P(U=k, V=n-j) = \binom{s}{t-j}\binom{n}{k} \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} \binom{im}{j}.$$

Example

Suppose we draw t balls without replacement from an urn containing m copies of each of n different solid colored balls and s identical striped balls.

$$\mathrm{E}\Big((C_0)_{(\nu)}(Y_1)_{(\tau)}\Big) = \frac{s_{(\tau)}n_{(\nu)}(s{+}m(n-v)-r)_{(t-r)}}{(mn+s)_{(t)}t_{(\tau)}}$$

Proof

$\mathrm{E}\Big((C_0)_{(\nu)}(Y_1)_{(\tau)}\Big) =$

$$= \frac{1}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\Big((1+\lambda)^s e^{\theta(1+\lambda)^m}\mathrm{E}\Big((W_0)_{(\nu)}(Z_1)_{(\tau)}\Big)\Big)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{1}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\Big((1+\lambda)^s e^{\theta(1+\lambda)^m}\mathrm{E}\Big((W_0)_{(\nu)}\Big)\mathrm{E}\Big((Z_1)_{(\tau)}\Big)\Big)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{1}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\big((1+\lambda)^s e^{\theta(1+\lambda)^m}(\theta^\nu)\big(\big(\tfrac{\lambda}{1+\lambda}\big)^\tau s_{(\tau)}\big)\big)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{s_{(\tau)}}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\big((1+\lambda)^s\big(\tfrac{\lambda}{1+\lambda}\big)^\tau\frac{d^n}{d\theta^n}\big(e^{\theta(1+\lambda)^m}\theta^\nu\big)\big)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{s_{(\tau)}\frac{n!}{(n-\nu)!}}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\Big((1+\lambda)^{s+m(n-v)-\tau}\lambda^\tau\Big)\Big|_{\lambda=0}$$

$$= \frac{s_{(\tau)}\frac{n!}{(n-\nu)!}}{\binom{mn+s}{t}t!}\binom{s+m(n-v)-\tau}{t-\tau}t!$$

$$= \frac{s_{(\tau)}\frac{n!}{(n-\nu)!}}{\binom{mn+s}{t}}\binom{s+m(n-v)-\tau}{t-\tau}$$

$$= \frac{s_{(\tau)}n_{(\nu)}}{\binom{mn+s}{t}}\binom{s+m(n-v)-\tau}{t-\tau}$$

$$= \frac{s_{(\tau)}n_{(\nu)}(s{+}m(n{-}v){-}\tau)_{(t-\tau)}(t-\tau)!}{(mn{+}s)_{(t)}t!}$$

$$= \frac{s_{(\tau)}n_{(\nu)}(s{+}m(n{-}v){-}r)_{(t-r)}}{(mn{+}s)_{(t)}t_{(\tau)}}$$

$$\mathrm{E}\Big((C_0)_{(\nu)}\Big) =$$

$$= \frac{1}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\Big((1+\lambda)^s e^{\theta(1+\lambda)^m}\mathrm{E}\Big((W_0)_{(\nu)}\Big)\Big)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{1}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\left((1+\lambda)^s e^{\theta(1+\lambda)^m}\theta^{\nu}\right)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{1}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\left((1+\lambda)^s \frac{d^n}{d\theta^n}\left(e^{\theta(1+\lambda)^m}\theta^{\nu}\right)\right)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{1}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\Big((1+\lambda)^s(1+\lambda)^{m(n-\nu)}\frac{n!}{(n-\nu)!}\Big)\Big|_{\lambda=0}$$

$$= \frac{\frac{n!}{(n-\nu)!}}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\Big((1+\lambda)^{s+m(n-v)}\Big)\Big|_{\lambda=0}$$

$$= \frac{\frac{n!}{(n-\nu)!}}{\binom{mn+s}{t}t!}\binom{s+m(n-v)}{t}t!$$

$$= \frac{\frac{n!}{(n-\nu)!}}{\binom{mn+s}{t}}\binom{s+m(n-v)}{t}$$

$$= \frac{n_{(v)}}{(mn+s)_{(t)}}(s+m(n-v))_{(t)}$$

$$\mathrm{E}\Big((Y_1)_{(\tau)}\Big) =$$

$$= \frac{1}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\Big((1+\lambda)^s e^{\theta(1+\lambda)^m}\mathrm{E}\Big((Z_1)_{(\tau)}\Big)\Big)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{1}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\frac{d^n}{d\theta^n}\left((1+\lambda)^s e^{\theta(1+\lambda)^m}\left(\left(\frac{\lambda}{1+\lambda}\right)^{\tau}s_{(\tau)}\right)\right)\Big|_{\substack{\lambda=0\\ \theta=0}}$$

$$= \frac{1}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\left((1+\lambda)^s\left(\left(\frac{\lambda}{1+\lambda}\right)^{\tau}s_{(\tau)}\right)(1+\lambda)^{mn}\right)\Big|_{\lambda=0}$$

$$= \frac{s_{(\tau)}}{\binom{mn+s}{t}t!}\frac{d^t}{d\lambda^t}\left((1+\lambda)^{s-\tau+mn}\lambda^{\tau}\right)\Big|_{\lambda=0}$$

$$= \frac{s_{(\tau)}}{\binom{mn+s}{t}t!}\binom{s-\tau+mn}{t-\tau}t!$$

$$= \frac{s_{(\tau)}}{\binom{mn+s}{t}} \binom{s-\tau+mn}{t-\tau}$$