Sampling Designs on Finite Populations with Spreading Control Parameters

Yves Tillé, Lionel Qualité and Matthieu Wilhelm

University of Neuchâtel,

Swiss Federal Office of Statistics and University of Neuchâtel,

University of Neuchâtel,

Supplementary Material

Appendix A: Proof of Proposition ?? and Remark ??

Lemma 1. *If $f(\cdot)$ is a probability distribution on $\{1,2,\dots,\}$ with cumulative distribution function $F(\cdot)$, and $k, j \geq 1$, then*

$$\sum_{t=1}^{k} f^{(j+1)*}(t) = \sum_{t=1}^{k} f^{j*}(t) F(k-t).$$

Proof. Indeed, if $\mathbf{1}_A$ is the indicator function of set A,

$$\begin{aligned}
\sum_{t=1}^{k} f^{(j+1)*}(t) &= \sum_{t=1}^{k}\sum_{u=1}^{t} f^{j*}(u)f(t-u), \\
&= \sum_{t}\sum_{u} f^{j*}(u)f(t-u)\mathbf{1}_{\{1\le u\le t\}}\ \mathbf{1}_{\{1\le t\le k\}}, \\
&= \sum_{u}\sum_{t} f^{j*}(u)f(t-u)\mathbf{1}_{\{1\le u\le k\}}\ \mathbf{1}_{\{1\le u\le t\}}\ \mathbf{1}_{\{1\le t\le k\}}, \\
&= \sum_{u} f^{j*}(u)\mathbf{1}_{\{1\le u\le k\}}\left[\sum_{t} f(t-u)\ \mathbf{1}_{\{1\le u\le t\}}\mathbf{1}_{\{1\le t\le k\}}\right], \\
&= \sum_{u=1}^{k} f^{j*}(u)\left[F(k-u) - \underbrace{F(0)}_{=0}\right], \\
&= \sum_{t=1}^{k} f^{j*}(t)F(k-t).
\end{aligned}$$

□

Proof of Proposition **??**. $f_0(\cdot)$ is a well-defined non-negative function on $\mathbb{N}$.

It is sufficient to prove that $\sum_{k\ge0} f(\{k+1,\dots\}) = \mu$, but

$$\begin{aligned}
\sum_{k\ge0} f(\{k+1,\dots\}) &= \sum_{k\ge0}\sum_{j\ge k+1} f(j), \\
&= \sum_{j\ge0}\sum_{k\ge0} f(j)\mathbf{1}_{k+1\le j}, \\
&= \sum_{j\ge0} j\cdot f(j), \\
&= \mu.
\end{aligned}$$

As $f_0(k-t) = [1 - F(k-t)]/\mu$, to prove (**??**), it is sufficient to note that

$$\begin{aligned}
\sum_{t=1}^{k}[1-F(k-t)]\sum_{j=1}^{t} f^{j*}(t) &= \sum_{t}\sum_{j}[1-F(k-t)]\, f^{j*}(t)\mathbf{1}_{1\le t\le k}\mathbf{1}_{1\le j\le t}, \\
&= \sum_{j}\sum_{t}[1-F(k-t)]\, f^{j*}(t)\mathbf{1}_{1\le t\le k}\mathbf{1}_{1\le j\le t}, \\
&= \sum_{j}\left[\sum_{t} f^{j*}(t)\mathbf{1}_{1\le t\le k}\mathbf{1}_{1\le j\le t} - \sum_{t} F(k-t)f^{j*}(t)\mathbf{1}_{1\le t\le k}\mathbf{1}_{1\le j\le t}\right], \\
&= \sum_{j=1}^{k}\left[\sum_{t=j}^{k} f^{j*}(t) - \sum_{t=j}^{k} F(k-t)f^{j*}(t)\right], \\
&= \sum_{j=1}^{k}\left[\sum_{t=1}^{k} f^{j*}(t) - \sum_{t=1}^{k} F(k-t)f^{j*}(t)\right] \text{ (indeed, } f^{j*}(t)=0 \text{ if } t<j), \\
&= \sum_{t=1}^{k} f^{1*}(t) - \sum_{t=1}^{k} F(k-t)f^{k*}(t) \text{ via lemma 1}, \\
&= F(k) - \sum_{t=1}^{k} f^{(k+1)*}(t) = F(k),
\end{aligned}$$

since $f^{(k+1)*}(t) = 0$ if $t \le k$, and the result follows immediately. □

Proof of Remark **??**. Consider X a random variable on $\mathbb{N}$ with finite moment of order $m+1$, $\mathrm{E}(X^{m+1})$, $m \ge 0$, and its forward transform X_F according to Definition **??**. Then we can write:

$$\begin{aligned}
\sum_{k\ge 0} k^m \Pr(X_F = k) &= \sum_{k\ge 0} k^m \frac{\Pr(X\ge k)}{\mathrm{E}(X+1)} = \sum_{k\ge 0}\sum_{i\ge k}\frac{k^m \Pr(X=i)}{\mathrm{E}(X+1)}, \\
&= \frac{1}{\mathrm{E}(X+1)}\sum_{i\ge 0}\sum_{k\ge 0}\mathbf{1}_{k\le i} k^m \Pr(X=i) = \frac{1}{\mathrm{E}(X+1)}\sum_{i\ge 0}\left(\sum_{k=0}^{i} k^m\right)\Pr(X=i), \\
&= \frac{\mathrm{E}[F_m(X)]}{\mathrm{E}(X+1)},
\end{aligned}$$

where $F_m(x) = \sum_{k=0}^{x} k^m$. □

Appendix B

Proposition 1. *The lines of matrix* $\mathbf{A}$ *with general term* a_{kt} *given at (**??**) all sum to* n*.*

Proof. We have

$$a_{kt} = \mathbf{1}_{t=k} + \mathbf{1}_{t<k} \sum_{j=1}^{k-t} f_j(k-t) + \mathbf{1}_{t>k} \sum_{j=1}^{N+k-t} f_j(N+k-t),$$

with $f_j(t) = 0$ if $j < t$, $t \leq 1$, $t > N$ or $j > n$. We also have that $f_n(N) = 1$ and $f_j(N) = 0$ if $j < n$. The conclusion follows from

$$\begin{aligned}
\sum_{t=1}^{N} \mathbf{1}_{t<k} \sum_{j=1}^{k-t} f_j(k-t) &= \sum_{t=1}^{N} \sum_{j=1}^{N} f_j(k-t) \mathbf{1}_{j \leq k-t} \mathbf{1}_{j \leq n}, \\
&= \sum_{j=1}^{n} \sum_{t=1}^{N} f_j(k-t) = \sum_{j=1}^{n} \Pr(S_j \leq k-1), \text{ and} \\
\sum_{t=1}^{N} \mathbf{1}_{t>k} \sum_{j=1}^{N+k-t} f_j(N+k-t) &= \sum_{t=1}^{N} \sum_{j=1}^{N} f_j(N+k-t) \mathbf{1}_{j \leq N+k-t} \mathbf{1}_{j \leq n} \mathbf{1}_{t>k}, \\
&= \sum_{j=1}^{n} \sum_{t=1}^{N} f_j(N+k-t) \mathbf{1}_{t>k} = \sum_{j=1}^{n} \left[\Pr(S_j \geq k) - f_j(N) \right].
\end{aligned}$$

□

Appendix C: discrete probability distributions

Let $\mathbb{R}_+$ denote the set of positive real numbers,

$$\Gamma(r,x) = \int_x^{+\infty} t^{r-1}e^{-t}\; dt, \quad \gamma(r,x) = \int_0^x t^{r-1}e^{-t}\; dt,$$

where $r > 0, x > 0$ and

$$\mathrm{B}_x(a,b) = \int_0^x t^{a-1}(1-t)^{b-1}dt, \qquad \mathrm{I}_x(a,b) = \frac{\mathrm{B}_x(a,b)}{\mathrm{B}(a,b)},$$

with $a > 0, b > 0, 0 < x < 1$.

Table 1: Discrete distributions of probability

Name	Notation	PMF	Support	Parameters	Mean	Variance
Bernoulli	$\mathcal{Bern}(p)$	$p^x(1-p)^{1-x}$	$\{0,1\}$	$p \in [0,1]$	p	$p(1-p)$
Forward Bernoulli	$\mathcal{ForBern}(p)$	$\frac{p^x}{p+1}$	$\{0,1\}$	$p \in [0,1], n \in \mathbb{N}$	(see below the table)	
Binomial	$\mathcal{Bin}(n,p)$	$\binom{n}{x}p^x(1-p)^{n-x}$	$\{0,\dots,n\}$	$p \in [0,1], n \in \mathbb{N}$	np	$np(1-p)$
Forward Binomial	$\mathcal{ForBin}(n,p)$	$\frac{\mathrm{I}_p(x,n-x+1)}{np+1}$	$\{0,\dots,n\}$	$p \in [0,1], n \in \mathbb{N}$	(see below the table)	
Geometric	$\mathcal{G}(1-p)$	$p(1-p)^x$	$\mathbb{N}$	$p \in [0,1]$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	$\mathcal{NB}(r,p)$	$\frac{\Gamma(r+x)}{x!\Gamma(r)}p^r(1-p)^x$	$\mathbb{N}$	$p \in [0,1], r \in \mathbb{N}^*$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
Forward Negative Binomial	$\mathcal{ForNB}(r,p)$	$\frac{p\mathrm{I}_{(1-p)}(x,r)}{r(1-p)+p}$	$\mathbb{N}$	$p \in [0,1], r \in \mathbb{N}^*$	(see below the table)	
Poisson	$\mathcal{P}(\lambda)$	$\frac{e^{-\lambda}\lambda^x}{x!}$	$\mathbb{N}$	$\lambda \in \mathbb{R}_+$	λ	λ
Forward Poisson	$\mathcal{ForP}(\lambda)$	$\frac{1}{\lambda+1}\left[\mathbf{1}_{x=0} + \frac{\gamma(x,\lambda)}{(x-1)!}\mathbf{1}_{x\geq 1}\right]$	$\mathbb{N}$	$\lambda \in \mathbb{R}_+$	(see below the table)	
Hypergeometric	$\mathcal{H}(m,r,R)$	$\frac{\binom{r}{x}\binom{R-r}{m-x}}{\binom{R}{m}}$	$\{0,\dots,m\} \bigcap \{r+m-R,\dots,r\}$	$m,r,R \in \mathbb{N}^*$, $m,r \leq R$	$\frac{mr}{R}$	$\frac{mr(R-r)}{R^2}\frac{R-m}{R-1}$
Negative Hypergeometric	$\mathcal{NH}(m,r,R)$	$\frac{\frac{\Gamma(r+x)}{\Gamma(r)x!}\frac{\Gamma(R-r+m-x)}{\Gamma(R-r)(m-x)!}}{\frac{\Gamma(m+R)}{\Gamma(R)m!}}$	$\{0,\dots,m\}$	$m,r,R \in \mathbb{N}^*$ $1 \leq R-r$	$\frac{mr}{R}$	$\frac{mr(R-r)}{R^2}\frac{R+m}{R+1}$
Uniform	$\mathcal{U}(0,a)$	$\frac{1}{a+1}$	$\{0,\dots,a\}$	$a \in \mathbb{N}$	$\frac{a}{2}$	$\frac{(a+1)^2-1}{12}$

Expectations and variances of forward distributions are easily computed in function of the first three moments of the original distribution (see Remark **??**).