Summationer

Ligefordelingen på $\{1, 2, \ldots, n\}$

Sandsynlighedsfunktion:

$$P(X = k) = \frac{1}{n}, \quad k = 1, 2, \ldots, n$$

Kontrol:

$$\sum_{k=1}^{n} \frac{1}{n} = \frac{1}{n} \sum_{k=1}^{n} 1 = \frac{1}{n} n = 1$$

Middelværdi og varians:

$$\mathrm{E}X = \sum_{k=1}^{n} k \frac{1}{n} \overset{1)}{=} \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}$$

$$\mathrm{E}[X^2] = \sum_{k=1}^{n} k^2 \frac{1}{n} \overset{2)}{=} \frac{1}{n} \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$$

$$\begin{aligned} \mathrm{Var}X = \mathrm{E}[X^2] - (\mathrm{E}X)^2 &= \frac{(n+1)(2n+1)}{6} - \Big(\frac{n+1}{2}\Big)^2 \\ &= (n+1)\frac{4n+2-3n-3}{12} = \frac{(n+1)(n-1)}{12} = \frac{n^2-1}{12} \end{aligned}$$

---

1) $\sum_{k=1}^{n} k = \sum_{k=1}^{n} (n+1-k) = \frac{1}{2} \sum_{k=1}^{n} (k+n+1-k) = \frac{1}{2} \sum_{k=1}^{n} (n+1) = \frac{n(n+1)}{2}$

2) $$\begin{aligned} \sum_{k=1}^{n} k^2 &= \sum_{k=1}^{n} \sum_{j=1}^{k} (j^2 - (j-1)^2) = \sum_{k=1}^{n} \sum_{j=1}^{k} (j + j - 1)(j - (j-1)) = \sum_{k=1}^{n} \sum_{j=1}^{k} (2j-1) \\ &= \sum_{j=1}^{n} (2j-1) \sum_{k=j}^{n} 1 = \sum_{j=1}^{n} (2j-1)(n-(j-1)) = \sum_{j=1}^{n} (2nj - 2j^2 + 2j - n + j - 1) \\ &= -2\sum_{j=1}^{n} j^2 + (2n+3)\sum_{j=1}^{n} j - \sum_{j=1}^{n} (n+1) = -2\sum_{j=1}^{n} j^2 + (2n+3)\frac{n(n+1)}{2} - n(n+1) \\ &\Rightarrow \quad 3\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+3-2)}{2} \quad \Rightarrow \quad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} \end{aligned}$$

Binomialfordelingen, $X \sim b(n, p)$

Sandsynlighedsfunktion:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

Kontrol:

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

Middelværdi og varians:

$$\begin{aligned}
\mathrm{E}X &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= p \sum_{k=1}^{n} n \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} = np,
\end{aligned}$$

idet den sidste sum er summen af alle punktsandsynligheder $P(Y = k)$, hvor $Y \sim b(n-1, p)$. Denne sum er derfor lig med 1.

$$\begin{aligned}
\mathrm{E}[X(X-1)] &= \sum_{k=0}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=2}^{n} k(k-1) \binom{n}{k} p^k (1-p)^{n-k} \\
&= p^2 \sum_{k=2}^{n} n(n-1) \binom{n-2}{k-2} p^{k-2} (1-p)^{(n-2)-(k-2)} \\
&= n(n-1)p^2 \sum_{k=0}^{n} \binom{n-2}{k} p^k (1-p)^{(n-2)-k} \\
&= n(n-1)p^2
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}X &= \mathrm{E}[X^2] - (\mathrm{E}X)^2 = \mathrm{E}[X(X-1)] + \mathrm{E}X - (\mathrm{E}X)^2 \\
&= n(n-1)p^2 + np + (np)^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np(1-p)
\end{aligned}$$

Note: Bemærk, at $X = X_1 + X_2 + \cdots + X_n$, hvor X_i'erne er uafhængige Bernoullifordelte variable, $X_i \sim b(1, p)$. Heraf fås en simplere udregning af middelværdi og varians:

$$\mathrm{E}X = \sum_{i=1}^{n} \mathrm{E}X_i = \sum_{i=1}^{n} p = np$$

$$\mathrm{Var}X = \sum_{i=1}^{n} \mathrm{Var}X_i = \sum_{i=1}^{n} p(1-p) = np(1-p)$$

Den hypergeometriske fordeling, $X \sim h(M,N,n)$

Sandsynlighedsfunktion:

$$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad k=\max\{0,n+M-N\},\dots,\min\{M,n\}$$

Ved benyttelse af konventionen $\binom{n}{k}=0$ for $k<0$ og for $k>n$ kan vi angive X's værdier til $0,1,\dots,n$.

Kontrol:

$$\sum_{k=0}^{n}\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}=\frac{1}{\binom{N}{n}}\sum_{k=0}^{n}\binom{M}{k}\binom{N-M}{n-k}=\frac{1}{\binom{N}{n}}\binom{N}{n}=1$$

Middelværdi og varians:

$$\begin{aligned}
\mathrm{E}X&=\sum_{k=0}^{n}k\,\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}=\sum_{k=1}^{n}\frac{k\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}\\
&=\sum_{k=1}^{n}\frac{M\binom{M-1}{k-1}\binom{(N-1)-(M-1)}{(n-1)-(k-1)}}{\frac{N}{n}\binom{N-1}{n-1}}\\
&=\frac{nM}{N}\sum_{k=0}^{n-1}\frac{\binom{M-1}{k}\binom{(N-1)-(M-1)}{(n-1)-k}}{\binom{N-1}{n-1}}\\
&=n\,\frac{M}{N}\,,
\end{aligned}$$

idet den sidste sum er summen af alle punktsandsynligheder $P(Y=k)$, hvor $Y\sim h(M-1,N-1,n-1)$.

$$\begin{aligned}
\mathrm{E}[X(X-1)]&=\sum_{k=0}^{n}k(k-1)\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}=\sum_{k=2}^{n}\frac{k(k-1)\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}\\
&=\sum_{k=2}^{n}\frac{M(M-1)\binom{M-2}{k-2}\binom{(N-2)-(M-2)}{(n-2)-(k-2)}}{\frac{N(N-1)}{n(n-1)}\binom{N-2}{n-2}}\\
&=\frac{n(n-1)M(M-1)}{N(N-1)}\sum_{k=0}^{n-2}\frac{\binom{M-2}{k}\binom{(N-2)-(M-2)}{(n-2)-k}}{\binom{N-2}{n-2}}
\end{aligned}$$

$$= \frac{n(n-1)M(M-1)}{N(N-1)}$$

$$\begin{aligned}
\mathrm{Var}X &= \mathrm{E}[X^2] - (\mathrm{E}X)^2 = \mathrm{E}[X(X-1)] + \mathrm{E}X - (\mathrm{E}X)^2 \\
&= \frac{n(n-1)M(M-1)}{N(N-1)} + \frac{nM}{N} - \Big(\frac{nM}{N}\Big)^2 \\
&= \frac{nM}{N}\Big(\frac{nM-n-M+1}{N-1} + 1 - \frac{nM}{N}\Big) \\
&= \frac{nM}{N}\,\frac{nMN - mN - MN + N + N^2 - N - nMN + nM}{N(N-1)} \\
&= \frac{nM}{N}\,\frac{N(N-n) - M(N-n)}{N(N-1)} \\
&= \frac{nM}{N}\,\frac{(N-M)(N-n)}{N(N-1)} \\
&= n\,\frac{M}{N}\Big(1-\frac{M}{N}\Big)\frac{N-n}{N-1}
\end{aligned}$$

Note: Ved benyttelse af indikatorvariable kan X skrives som $X = I_1 + I_2 + \cdots + I_n$, $\mathrm{E}I_i = \frac{M}{N}$, $\mathrm{Var}E_i = \frac{M}{N}\big(1-\frac{M}{N}\big)$, $\mathrm{Cov}(I_i, I_j) = \frac{M}{N}\frac{M-1}{N-1} - \big(\frac{M}{N}\big)^2$, $i \neq j$. Heraf fås en alternativ udregning af middelværdi og varians:

$$\mathrm{E}X = \sum_{i=1}^{n} \mathrm{E}I_i = n\,\frac{M}{N}$$

$$\begin{aligned}
\mathrm{Var}X &= \sum_{i=1}^{n} \mathrm{Var}I_i + \sum_{i\neq j} \mathrm{Cov}(I_i, I_j) \\
&= n\,\frac{M}{N}\Big(1-\frac{M}{N}\Big) + n(n-1)\Big(\frac{M}{N}\,\frac{M-1}{N-1} - \Big(\frac{M}{N}\Big)^2\Big) \\
&= n\,\frac{M}{N}\Big[\frac{N-M}{N} + (n-1)\frac{MN-N-MN+M}{N(N-1)}\Big] \\
&= n\,\frac{M}{N}\,\frac{N-M}{N}\Big(1 - \frac{n-1}{N-1}\Big) \\
&= n\,\frac{M}{N}\Big(1-\frac{M}{N}\Big)\frac{N-n}{N-1}
\end{aligned}$$

Poissonfordelingen, $X \sim p(\lambda)$

Sandsynlighedsfunktion:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, \ldots$$

Kontrol:

$$\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1$$

Middelværdi og varians:

$$\mathrm{E}X = \sum_{k=0}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!} = \lambda\sum_{k=1}^{\infty}\frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} = \lambda\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!} = \lambda$$

$$\mathrm{E}[X(X-1)] = \sum_{k=0}^{\infty} k(k-1)\frac{e^{-\lambda}\lambda^k}{k!} = \lambda^2\sum_{k=2}^{\infty}\frac{e^{-\lambda}\lambda^{k-2}}{(k-2)!} = \lambda^2\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!} = \lambda^2$$

$$\mathrm{Var}X = \mathrm{E}[X^2] - (\mathrm{E}X)^2 = \mathrm{E}[X(X-1)] + \mathrm{E}X - (\mathrm{E}X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Den geometriske fordeling, $X \sim g(p)$

Sandsynlighedsfunktion:

$$P(X=k) = p(1-p)^{k-1}, \quad k = 1, 2, \ldots$$

Kontrol:

$$\sum_{k=1}^{\infty} p(1-p)^{k-1} = p\sum_{k=0}^{\infty}(1-p)^k = p\frac{1}{1-(1-p)} = 1$$

Middelværdi og varians:

$$\mathrm{E}X = \sum_{k=0}^{\infty} P(X>k) \overset{3)}{=} \sum_{k=0}^{\infty}(1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p}$$

$$\begin{aligned}\mathrm{E}[X(X-1)] &= \sum_{k=0}^{\infty} 2kP(X>k) \overset{4)}{=} 2\sum_{k=0}^{\infty} k(1-p)^k \\ &= \frac{2(1-p)}{p}\sum_{k=1}^{\infty} kp(1-p)^{k-1} = \frac{2(1-p)}{p}\mathrm{E}X = \frac{2(1-p)}{p^2}\end{aligned}$$

$$\begin{aligned}\mathrm{Var}X &= \mathrm{E}[X^2] - (\mathrm{E}X)^2 = \mathrm{E}[X(X-1)] + \mathrm{E}X - (\mathrm{E}X)^2 \\ &= \frac{2(1-p)}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{2-2p+p-1}{p^2} = \frac{1-p}{p^2}\end{aligned}$$

Den stokastiske variabel $Y = X - 1$ siges også at være geometrisk fordelt. Sandsynlighedsfunktion, middelværdi og varians for Y er

$$P(Y=k) = p(1-p)^k, \quad k = 0, 1, \ldots$$

[3] Alternativt: $\mathrm{E}X = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p\sum_{k=1}^{\infty} k(1-p)^{k-1} = p\frac{1}{(1-(1-p))^2} = \frac{1}{p}$

[4] Alternativt: $\mathrm{E}[X(X-1)] = \sum_{k=1}^{\infty} k(k-1)p(1-p)^{k-1} = p(1-p)\sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2}$

$$= p(1-p)\frac{2}{(1-(1-p))^3} = \frac{2(1-p)}{p^2}$$

$$\mathrm{E}Y = \mathrm{E}(X-1) = \mathrm{E}X - 1 = \frac{1}{p} - 1 = \frac{1-p}{p}$$
$$\mathrm{Var}Y = \mathrm{Var}(X-1) = \mathrm{Var}X = \frac{1-p}{p^2}$$

Den negative binomialfordeling, $X \sim nb(r,p)$

Sandsynlighedsfunktion:

$$P(X=r) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \dots$$

Kontrol:

$$\begin{aligned}
\sum_{k=r}^{\infty} \binom{k-1}{r-1} p^r (1-p)^{k-r} &= p^r \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} (1-p)^k \\
&= p^r \sum_{k=0}^{\infty} \binom{k+r-1}{k} (1-p)^k \\
&= p^r \sum_{k=0}^{\infty} (-1)^k \binom{-r}{k} (1-p)^k \\
&= p^r \sum_{k=0}^{\infty} \binom{-r}{k} (p-1)^k \\
&= p^r (1+p-1)^{-r} = p^r p^{-r} = 1
\end{aligned}$$

I regningerne blev det benyttet, at

$$\begin{aligned}
\binom{k+r-1}{k} = \binom{r+k-1}{k} &= \frac{(r+k-1)(r+k-2)\cdots(r+1)r}{k!} \\
&= (-1)^k \frac{-r(-r-1)\cdots(-r-k+2)(-r-k+1)}{k!} \\
&= (-1)^k \binom{-r}{k}
\end{aligned}$$

Endvidere blev formlen i baggrundsnote til sandsynlighedsregning side 6 linie 3 fra neden anvendt.

Middelværdi og varians:

$$\begin{aligned}
\mathrm{E}X &= \sum_{k=r}^{\infty} k \binom{k-1}{r-1} p^r (1-p)^{k-r} \\
&= p^r \sum_{k=r}^{\infty} r \binom{k}{r} (1-p)^{k-r} = r p^r \sum_{k=0}^{\infty} \binom{k+r}{r} (1-p)^k \\
&= r p^r \sum_{k=0}^{\infty} \binom{k+r}{k} (1-p)^k \\
&= r p^r \sum_{k=0}^{\infty} (-1)^k \binom{-(r+1)}{k} (1-p)^k
\end{aligned}$$

$$
\begin{aligned}
&= rp^r \sum_{k=0}^{\infty} \binom{-(r+1)}{k} (p-1)^k \\
&= \frac{r}{p} p^{r+1} (1+p-1)^{-(r+1)} = \frac{r}{p}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}[X(X+1)] &= \sum_{k=r}^{\infty} k(k+1) \binom{k-1}{r-1} p^r (1-p)^{k-r} \\
&= p^r \sum_{k=r}^{\infty} r(r+1) \binom{k+1}{r+1} (1-p)^{k-r} \\
&= r(r+1)p^r \sum_{k=0}^{\infty} \binom{k+r+1}{r+1} (1-p)^k \\
&= r(r+1)p^r \sum_{k=0}^{\infty} \binom{k+r+1}{k} (1-p)^k \\
&= r(r+1)p^r \sum_{k=0}^{\infty} (-1)^k \binom{-(r+2)}{k} (1-p)^k \\
&= r(r+1)p^r \sum_{k=0}^{\infty} \binom{-(r+2)}{k} (p-1)^k \\
&= \frac{r(r+1)}{p^2} p^{r+2} (1+p-1)^{-(r+2)} = \frac{r(r+1)}{p^2}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}X &= \mathrm{E}[X^2] - (\mathrm{E}X)^2 = \mathrm{E}[X(X+1)] - \mathrm{E}X - (\mathrm{E}X)^2 \\
&= \frac{r(r+1)}{p^2} - \frac{r}{p} - \left(\frac{r}{p}\right)^2 = \frac{r^2 + r - rp - r^2}{p^2} \\
&= \frac{r(1-p)}{p^2}
\end{aligned}
$$

Note: Bemærk, at $X = X_1 + X_2 + \cdots + X_r$, hvor X_i'erne er uafhængige geometrisk fordelte variable, $X_i \sim g(p)$. Heraf fås en simplere udregning af middelværdi og varians:

$$
\begin{aligned}
\mathrm{E}X &= \sum_{i=1}^{r} \mathrm{E}X_i = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p} \\
\mathrm{Var}X &= \sum_{i=1}^{r} \mathrm{Var}X_i = \sum_{i=1}^{r} \frac{1-p}{p^2} = \frac{r(1-p)}{p^2}
\end{aligned}
$$

1.3.2013/BR