Ordensvariable

Definition

Betragt uafhængige og identisk fordelte kontinuerte stokastiske variable $X_1, X_2, \ldots, X_n$ med fordelingsfunktion $F(x)$ og tæthedsfunktion $f(x)$.

Når $X_1, X_2, \ldots, X_n$ arrangeres efter størrelse, fremkommer de såkaldte ordensvariable $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$. Der gælder altså, at

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}.$$

Bemærk, at de ordensvariable er afhængige.

Simultan fordelingsfunktion for $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$

I henhold til definitionen af fordelingsfunktion har vi

$$F_{(X_{(1)},X_{(2)},\ldots,X_{(n)})}(x_1, x_2, \ldots, x_n) = P(X_{(1)} \leq x_1, X_{(2)} \leq x_2, \ldots, X_{(n)} \leq x_n),$$
$$-\infty < x_1 < x_2 < \cdots < x_n < \infty.$$

Lad π betegne en permutation af mængden $\{1, 2, \ldots, n\}$, dvs $\pi \in (S_n, \circ)$, den symmetriske gruppe af n'te orden. Bemærk, at hændelserne

$$\{-\infty < X_{\pi(1)} \leq x_1, x_1 < X_{\pi(2)} \leq x_2, \ldots, x_{n-1} < X_{\pi(n)} \leq x_n\}, \quad \pi \in S_n,$$

er disjunkte, hvoraf følger, at

$$F_{(X_{(1)},X_{(2)},\ldots,X_{(n)})}(x_1, x_2, \ldots, x_n)$$
$$= \sum_{\pi \in S_n} P(-\infty < X_{\pi(1)} \leq x_1, x_1 < X_{\pi(2)} \leq x_2, \ldots x_{n-1} < X_{\pi(n)} \leq x_n),$$
$$-\infty < x_1 < x_2 < \cdots < x_n < \infty.$$

Af symmetrigrunde må alle sandsynlighederne, der indgår i denne sum, være ens, hvorefter

$$F_{(X_{(1)},X_{(2)},\ldots,X_{(n)})}(x_1, x_2, \ldots, x_n)$$
$$= n!\, P(-\infty < X_1 \leq x_1, x_1 < X_2 \leq x_2, \ldots, x_{n-1} < X_n \leq x_n),$$
$$-\infty < x_1 < x_2 < \cdots < x_n < \infty.$$

En nærmere udregning af denne sandsynlighed er særdeles kompliceret. For $F_{(X_{(1)},X_{(n)})}(x_1, x_n)$ se dog side 3-4.

Da de stokastiske variable $X_1, X_2, \ldots, X_n$ er uafhængige og identisk fordelte, og den simultane tæthedsfunktion for $(X_, X_2, \ldots, X_n)$ derfor har formen

$$f_{(X_1,X_2,\ldots,X_n)}(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\cdots f(x_n), \quad x_i \in \mathbb{R}, \quad i = 1, \ldots, n,$$

kan den simultane fordelingsfunktion for $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ alternativt skrives

$$\begin{aligned}
&F_{(X_{(1)},X_{(2)},\ldots,X_{(n)})}(x_1, x_2, \ldots, x_n) \\
&\qquad = n! \int_{\{(x_1,x_2,\ldots,x_n) | -\infty < x_1 < x_2 < \cdots < x_n < \infty\}} f(t_1)f(t_2)\cdots f(t_n) d\Omega, \\
&\qquad\qquad -\infty < x_1 < x_2 < \cdots < x_n < \infty.
\end{aligned}$$

Simultan tæthedsfunktion for $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$

Den simultane tæthedsfunktion fås umiddelbart ved differentiation af fordelingsfunktionen:

$$\begin{aligned}
f_{(X_{(1)},X_{(2)},\ldots,X_{(n)})}(x_1, x_2, \ldots, x_n) &= \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F_{(X_{(1)},X_{(2)},\ldots,X_{(n)})}(x_1, x_2, \ldots, x_n) \\
&= n!\, f(x_1)f(x_2)\cdots f(x_n), \\
&\qquad -\infty < x_1 < x_2 < \cdots < x_n < \infty
\end{aligned}$$

Øvelse. Lad D betegne definitionsmængden for den simultane tæthedsfunktion, og bemærk, at $D \subseteq \mathbb{R}^n$. Eftervis, at

$$\int_D f_{(X_{(1)},X_{(2)},\ldots,X_{(n)})}(x_1, x_2, \ldots, x_n)\, \mathrm{d}\Omega = 1.$$

Vink: Bestem successivt de marginale tæthedsfunktioner for $(X_{(2)}, \ldots, X_{(n)})$, $(X_{(3)}, \ldots, X_{(n)}), \ldots, X_{(n)}$. □

Fordelingsfunktion for den k'te mindste ordensvariabel $X_{(k)}$

Lad N være en stokastisk variabel, der angiver antallet af hændelser $\{X_j \leq x\}$, der fremkommer blandt de n mulige. Der gælder, at $N \sim b(n, F(x))$, og endvidere at $\{X_{(k)} \leq x\} = \{N \geq k\}$. Med benyttelse af N kan fordelingsfunktionen for $X_{(k)}$ opstilles:

$$\begin{aligned}
F_{X_{(k)}}(x) = P(X_{(k)} \leq x) &= \sum_{j=k}^{n} P(N = k) \\
&= \sum_{j=k}^{n} \binom{n}{j} \big(F(x)\big)^j \big(1 - F(x)\big)^{n-j}, \quad -\infty < x < \infty
\end{aligned}$$

Specielt for $X_{(1)}$ og $X_{(n)}$:

$$F_{X_{(1)}}(x) = \sum_{j=1}^{n} \binom{n}{j} \big(F(x)\big)^j \big(1 - F(x)\big)^{n-j}$$

$$
\begin{aligned}
&= 1 - \binom{n}{0}\big(F(x)\big)^0\big(1-F(x)\big)^{n-0} \\
&= 1 - \big(1-F(x)\big)^n, \qquad -\infty < x < \infty
\end{aligned}
$$

$$
\begin{aligned}
F_{X_{(n)}}(x) &= \binom{n}{n}\big(F(x)\big)^n\big(1-F(x)\big)^{n-n} \\
&= \big(F(x)\big)^n, \qquad -\infty < x < \infty
\end{aligned}
$$

Tæthedsfunktion for den k'te mindste ordensvariabel $X_{(k)}$

Tæthedsfunktionen for $X_{(k)}$ fås ved differentiation af fordelingsfunktionen:

$$
\begin{aligned}
f_{X_{(k)}}(x) &= \sum_{j=k}^{n}\binom{n}{j}f(x)\Big(j\big(F(x)\big)^{j-1}\big(1-F(x)\big)^{n-j} - (n-j)\big(F(x)\big)^{j}\big(1-F(x)\big)^{n-j-1}\Big) \\
&= \sum_{j=k}^{n} j\binom{n}{j}f(x)\big(F(x)\big)^{j-1}\big(1-F(x)\big)^{n-j} \\
&\qquad\qquad - \sum_{j=k+1}^{n}(n-j+1)\binom{n}{j-1}f(x)\big(F(x)\big)^{j-1}\big(1-F(x)\big)^{n-j}
\end{aligned}
$$

Idet der gælder, at $(n-j+1)\binom{n}{j-1} = j\binom{n}{j}$, går alle leddene på nær ét ud mod hinanden, hvorefter

$$
f_{X_{(k)}}(x) = k\binom{n}{k}f(x)\big(F(x)\big)^{k-1}\big(1-F(x)\big)^{n-k}, \quad -\infty < x < \infty.
$$

Specielt for $X_{(1)}$ og $X_{(n)}$:

$$
\begin{aligned}
f_{X_{(1)}}(x) &= 1\binom{n}{1}f(x)\big(F(x)\big)^0\big(1-F(x)\big)^{n-1} \\
&= nf(x)\big(1-F(x)\big)^{n-1}, \quad -\infty < x < \infty
\end{aligned}
$$

$$
\begin{aligned}
f_{X_{(n)}}(x) &= n\binom{n}{n}f(x)\big(F(x)\big)^{n-1}\big(1-F(x)\big)^0 \\
&= nf(x)\big(F(x)\big)^{n-1}, \quad -\infty < x < \infty
\end{aligned}
$$

Øvelse. Udled tæthedsfunktionen for $X_{(k)}$ ud fra den simultane tæthedsfunktion for $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$. Vink: Bestem først de marginale tæthedsfunktioner for $(X_{(2)}, \ldots, X_{(n)}), \ldots, (X_{(k)}, \ldots, X_{(n)})$, og dernæst de marginale tæthedsfunktioner for $(X_{(k)}, \ldots, X_{(n-1)}), \ldots, X_{(k)}$. □

Simultan fordelingsfunktion og tæthedsfunktion for $(X_{(1)}, X_{(n)})$

Fordelingsfunktionen:

$$
F_{(X_{(1)},X_{(n)})}(x_1, x_n) = P\big(X_{(1)} \le x_1, X_{(n)} \le x_n\big)
$$

$$
\begin{aligned}
&= P\big(X_{(n)} \le x_n\big) - P\big(X_{(1)} > x_1, X_{(n)} \le x_n\big) \\
&= F_{X_{(n)}}(x_n) - \prod_{j=1}^{n}\big(F_{X_j}(x_n) - F_{X_j}(x_1)\big) \\
&= \big(F(x_n)\big)^n - \big(F(x_n) - F(x_1)\big)^n, \quad -\infty < x_1 < x_n < \infty
\end{aligned}
$$

Undervejs blev det benyttet, at

$$\big\{X_{(1)} > x_1\big\} \cap \big\{X_{(n)} \le x_n\big\} = \bigcap_{j=1}^{n}\{x_1 < X_j \le x_n\}.$$

Tæthedsfunktionen:

$$
\begin{aligned}
f_{(X_{(1)},X_{(n)})}(x_1,x_n) &= \frac{\partial^2}{\partial x_1 \partial x_n} F_{(X_{(1)},X_{(n)})}(x_1,x_n) \\
&= n(n-1) f(x_1) f(x_n)\big(F(x_n) - F(x_1)\big)^{n-2}, \quad -\infty < x_1 < x_n < \infty
\end{aligned}
$$

Tæthedsfunktion for variationsbredden R

Variationsbredden (*eng.* range) defineres som $R = X_{(n)} - X_{(1)}$. Betragt den todimensionale variabel $(X_{(1)}, R)$, der fremkommer af $(X_{(1)}, X_{(n)})$ ved variabelskiftet

$$\left.\begin{aligned} x_1 &= x_1 \\ r &= x_n - x_1 \end{aligned}\right\} \Leftrightarrow \left\{\begin{aligned} x_1 &= x_1 \\ x_n &= r + x_1 \end{aligned}\right., \quad J(x_1, r) = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1.$$

Tæthedsfunktionen for $(X_{(1)}, R)$ bliver

$$f_{(X_{(1)},R)}(x_1, r) = n(n-1) f(x_1) f(r + x_1)\big(F(r + x_1) - F(x_1)\big)^{n-2} |1|.$$

Heraf den marginale tæthedsfunktion for R:

$$f_R(r) = n(n-1) \int_{-\infty}^{\infty} f(u) f(r+u)\big(F(r+u) - F(u)\big)^{n-2}\, \mathrm{d}u, \quad 0 \le r < \infty$$

Eksempler

Eksempel 1. Betragt $X_i \sim U[0;1]$, $i = 1, \dots, n$, X_i'erne uafhængige.
Den simultane tæthedsfunktion for $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$:

$$f_{(X_{(1)},X_{(2)},\dots,X_{(n)})}(x_1, x_2, \dots, x_n) = n!\,, \qquad 0 \le x_1 < x_2 < \cdots < x_n \le 1,$$

dvs. $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ er ligefordelt på

$$\{(x_1, x_2, \dots, x_n) \in \mathbb{R}^n \mid 0 \le x_1 < x_2 < \cdots < x_n \le 1\}.$$

Tæthedsfunktionen for $X_{(k)}$:

$$
\begin{aligned}
f_{X_{(k)}}(x) &= k\binom{n}{k} x^{k-1}(1-x)^{n-k}, \qquad 0 \le x \le 1 \\
&= \frac{\Gamma(k + (n-k+1))}{\Gamma(k)\,\Gamma(n-k+1)} x^{k-1}(1-x)^{(n-k+1)-1}, \qquad 0 \le x \le 1,
\end{aligned}
$$

dvs. $X_{(k)}$ er betafordelt[1)] med parametrene k og $n-k+1$. Heraf

$$\mathrm{E}X_{(k)} = \frac{k}{n+1} \quad \text{og} \quad \mathrm{Var}X_{(k)} = \frac{k(n-k+1)}{(n+1)^2(n+2)}.$$

Tæthedsfunktionen for R:

$$\begin{aligned} f_R(r) &= n(n-1)\int_0^{1-r} r^{n-2}\,\mathrm{d}u \\ &= n(n-1)r^{n-2}(1-r), \qquad 0 \le r \le 1 \end{aligned}$$

□

Eksempel 2. Betragt $X_i \sim e(\lambda)$, $i = 1, \ldots, n$, X_i'erne uafhængige.
Den simultane tæthedsfunktion for $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$:

$$f_{(X_{(1)},X_{(2)},\ldots,X_{(n)})}(x_1, x_2, \ldots, x_n) = n!\,\lambda^n e^{-\lambda(x_1+x_2+\cdots+x_n)}, \qquad 0 \le x_1 < x_2 < \cdots < x_n < \infty$$

Tæthedsfunktionen for $X_{(k)}$:

$$f_{X_{(k)}}(x) = k\binom{n}{k}\lambda e^{-(n-k+1)\lambda x}\big(1-e^{-\lambda x}\big)^{k-1}, \qquad 0 \le x < \infty$$

For $X_{(1)}$:

$$f_{X_{(1)}}(x) = n\lambda e^{-n\lambda x}, \qquad 0 \le x < \infty, \qquad \text{dvs. } X_{(1)} \sim e(n\lambda)$$

For $X_{(n)}$:

$$f_{X_{(n)}}(x) = n\lambda e^{-\lambda x}\big(1-e^{-\lambda x}\big)^{n-1}, \qquad 0 \le x < \infty$$

Tæthedsfunktionen for R:

$$\begin{aligned} f_R(r) &= n(n-1)\int_0^{\infty} \lambda e^{-\lambda u}\lambda e^{-\lambda(r+u)}\big(e^{-\lambda u} - e^{-\lambda(r+u)}\big)^{n-2}\,\mathrm{d}u \\ &= (n-1)\,\lambda\, e^{-\lambda r}\big(1-e^{-\lambda r}\big)^{n-2}\int_0^{\infty} n\lambda e^{-n\lambda u}\,\mathrm{d}u \\ &= (n-1)\,\lambda\, e^{-\lambda r}\big(1-e^{-\lambda r}\big)^{n-2}, \qquad 0 \le r < \infty \end{aligned}$$

□

Øvelse. Kontroller, at alle tæthedsfunktioner i eksempel 1 og eksempel 2 integrerer op til 1. □

Øvelse. Vis, at i eksempel 1 bliver $\mathrm{E}R = \dfrac{n-1}{n+1}$ og $\mathrm{Var}R = \dfrac{2(n-1)}{(n+1)^2(n+2)}$. □

Øvelse. Udled middelværdi og varians i betafordelingen. □

23.6.2013/BR

[1)] En betafordelt stokastisk variabel har tæthedsfunktion $f(x) = \frac{1}{B(\alpha,\beta)}\,x^{\alpha-1}(1-x)^{\beta-1}$, $0 \le x \le 1$. Det kan vises, at $B(\alpha,\beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$, og at middelværdi og varians er hhv. $\frac{\alpha}{\alpha+\beta}$ og $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.