Poissonprosesser og levetidsfordelinger

Poissonfordeling som grensetilfelle for binomisk fordeling

La X være binomisk fordelt med fordeling

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \text{ for } x = 0, 1, \ldots n. \tag{1}$$

Forventningsverdien til X er $\lambda = np$. Vi skal studere hva som skjer hvis vi endrer parametrene i den binomiske fordelingen ved å la n vokse mot uendelig og p avta mot null slik at forventningen $\lambda = np$ holdes konstant. Matematisk gjør vi dette ved å la n vokse mot uendelig og sette inn $p = \lambda/n$. Variansen til fordelingen kan da skrives

$$\text{var}(X) = n(\lambda/n)(1 - \lambda/n) \tag{2}$$

og vi ser da at var(X) nærmer seg λ når n vokser mot uendelig. I grensen må vi derfor finne en fordeling som har både forventning og varians lik λ.

Likning (1) kan skrives på formen

$$P(X = x) = \frac{n(n-1)\ldots(n-x+1)}{1 \cdot 2 \cdots x} \frac{\lambda^x}{n^x} (1 - \lambda/n)^n (1 - \lambda/n)^{-x} \tag{3}$$

Vi ser her at

$$\frac{n(n-1)\ldots(n-x+1)}{n^x} = \frac{n}{n} \frac{n-1}{n} \cdots \frac{n-x+1}{n} \tag{4}$$

nærmer seg 1 når n vokser fordi hver faktor nærmer seg 1. Den siste faktoren i (3), $(1-\lambda/n)^{-x}$, nærmer seg også $1^{-x} = 1$. Det er videre kjent fra matematikken at nest siste faktoren $(1-\lambda/n)^n$ nærmer seg $e^{-\lambda}$. Dermed har vi at for enhver x vil sannsynligheten for at X tar denne verdien nærme seg

$$P(X=x) = \frac{\lambda^x}{x!}e^{-\lambda}, \text{ for } x = 0, 1, \ldots \tag{5}$$

som er poissonfordelingen med parameter λ som altså både er forventningen og variansen i fordelingen,

$$\mathrm{E}X = \mathrm{var}(X) = \lambda. \tag{6}$$

Poissons punktprosess

La nå $\lambda(u)$ være en vilkårlig ikke-negativ funksjon som vi skal kalle intensiteten til Poissonprosessen. Vi kan tenke oss at u er tiden og at vi studerer forekomsten av en type hendelser som opptrer på forskjellige tidspunkter. Anta at sannsynligheten for at det skjer en slik hendelse i et lite tidsintervall $(u, u+\Delta u)$ (egentlig i grensen når Δu går mot null) er lik $\lambda(u)\Delta u$, altså både proporsjonal med intervallets lengde og intensiteten $\lambda(u)$ som gjelder i dette tidsintervallet. Anta videre at prosessen ikke har noen hukommelse, dvs. at sannsynligheten for en hendelse i $(u, u+\Delta u)$ ikke avhenger av forhistorien til prosessen. Hendelser på alle ikke-overlappende delintervaller blir da uavhengige.

La nå X betegne antall hendelser i intervallet $(0,t)$ og innfør $\Lambda = \int_0^t \lambda(u)du$, som da er arealet under intensitetskurven $\lambda(u)$ fra tid 0 til t. Vi deler så intervallet $(0,t)$ inn i n intervaller definert ved tidspunktene $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n = t$ som er valgt slik at integralene over alle intervallene er like store, dvs

$$\int_{t_i}^{t_{i+1}} \lambda(u)du = \Lambda/n. \tag{7}$$

Når n velges veldig stor (går mot uendelig) vil da hvert intervall utgjøre et deleksperiment der sannsynligheten for at det skal skje en hendelse er $\lambda(t_i)(t_{i+1} - t_i)$ som i grensen er lik $\int \lambda(u)du$ over intervallet, som igjen er valgt lik Λ/n. Dermed vil X i grensen når n vokser mot uendelig ha den binomiske fordelingen

$$P(X = x) = \binom{n}{x}(\frac{\Lambda}{n})^x(1 - \Lambda/n)^{n-x}. \tag{8}$$

Vi har ovenfor vist at grensen for denne fordelingen når n vokser mot uendelig (lengden av intervallene går mot null) er Poissonfordelingen

$$P(X = x) = \frac{\Lambda^x}{x!}e^{-\Lambda}. \tag{9}$$

Vi har dermed vist at antall punkter i et vilkårlig tidsintervall (a, b) med de antakelsen vi har gjort vil være Poissonfordelt med parameter $\int_a^b \lambda(u)du$.

Hvis $\lambda(u)$ er konstant lik λ kalles denne prosessen en homogen Poissonprosess (eller bare en Poissonprosess), og ellers kaller vi den en inhomogen Poissonprosess.

Poissonprosess og levetid

Denne prosessen kan brukes til å studere levetidsfordelinger. Vi definerer tiden lik null ved det tidspunkt et individ er født og lar T betegne individets levetid. Teorien er generell og kan brukes til å studere levetidsfordelinger for både individer og f.eks. teknologiske komponenter. Vi definerer dødsraten $\lambda(t)$ ved alder t som grensen for sannsynligheten for at individet dør i intervallet $(t, t + \Delta t)$ gitt at det lever ved tid t dividert med Δt når Δt nærmer seg null. Matematisk kan dette uttrykkes som

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(T < t + \Delta t | T > t)}{\Delta t}. \tag{10}$$

Vi har da at sannsynligheten for at individet dør i intervallet $(t, t + \Delta t)$ gitt at det lever ved tid t er lik $\lambda(t)\Delta t$ i grensen når Δt går not null. Prosessen fram til individet dør er derfor en inhomogen Poissonprosess med intensitet $\lambda(t)$ og den første hendelsen i prosessen representerer dødstidspunktet.

Vi kan utnytter dette til å finne fordelingen til levetiden T uttrykt ved intensiteten i Poissonprosessen som altså også representerer dødsraten til individet.

La t nå være et vilkårlig tidspunkt etter individets fødsel og la X være antall hendelser i Poissonprosessen definert ved $\lambda(u)$ i intervallet $(0, t)$. Ifølge teorien ovenfor er da X poissonfordelt med parameter $\Lambda = \int_0^t \lambda(u)du$. Da gjelder likheten

$$\{T \geq t\} = \{X = 0\}. \tag{11}$$

Men sannsynligheten for hendelsen $X = 0$ er kjent fordi X er poissonfordelt med parameter Λ, dvs.

$$P(T \geq t) = P(X = 0) = \frac{\Lambda^0}{0!}e^{-\Lambda} = e^{-\Lambda}. \tag{12}$$

Dermed har vi også utledet den kumulative fordelingen til T,

$$F_T(t) = P(T \leq t) = 1 - e^{-\int_0^t \lambda(u)du}. \tag{13}$$

Ved å derivere (og bruke kjerneregelen) finner vi da sannsynlighetstettheten

$$f_T(t) = \lambda(t)e^{-\int_0^t \lambda(u)du}. \tag{14}$$

Vi har her brukt at den deriverte av integralet m.h.p. øvre grense er lik integranden $\lambda(u)$ insatt den øvre integrasjonsgrensen, dvs. $u = t$.

Likning (14) gir uttrykker den generelle sammenhengen mellom dødsrate og levetid ved at den gir fordelingen uttrykt ved dødsraten. Den motsatte sammenhengen, altså dødsraten uttrykt ved levetidsfordelingen, kan vi finne ved å skrive likning (13) på formen

$$\int_0^t \lambda(u)du = -\ln[1 - F_T(t)] \tag{15}$$

og derivere på begge sider m.h.p. t, som gir

$$\lambda(t) = \frac{f_T(t)}{1 - F_T(t)}. \tag{16}$$

Alternativt kan denne sammenhengen finnes direkte fra definisjonen av dødsrate gitt i likning (10). Vi ser at

$$\frac{P(T < t + \Delta t | T > t)}{\Delta t} = \frac{P(t < T < t + \Delta t)}{\Delta t \; P(T > t)} = \frac{F(t + \Delta t) - F(t)}{\Delta t[1 - F(t)]} \tag{17}$$

og siden $\lim_{\Delta t \to 0}[F(t + \Delta t) - F(t)]/\Delta t = \frac{dF(t)}{dt} = f(t)$ pr. definisjon av den deriverte, følger (16) direkte.

Eksempler på levetidsfordelinger

Eksempel a - Eksponensialfordelingen

Hvis dødsratren er konstant lik λ blir $\Lambda = \lambda t$ slik at den tilhørende Poisson-prosessen er homogen. Innsatt i (14) gir dette levetidsfordelingen

$$f_T(t) = \lambda e^{-\lambda t} \tag{18}$$

definert for $t \geq 0$, som er eksponensialfordelingen med parameter λ. At dødsraten er konstant betyr at individene ikke aldres ved at alle individer uansett alder har samme sannsynlighet for å dø i neste tidsintervall. Denne fordelingen har som kjent forventningsverdi $1/\lambda$ slik at forventet levetid er omvendt proporsjonal med dødsraten.

Eksempel b - Weibullfordelingen

En fordeling som brukes mye i levetidsanalyse er weibullfordelingen som har kumulativ fordeling på formen

$$F_T(t) = 1 - e^{-\alpha t^{\beta}} \tag{19}$$

og sannsynlighetstetthet $f_T(t) = \alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}}$. Ved å sette dett inn i (16) finner vi at den tilsvarende dødsraten må være $\lambda(t) = \alpha\beta t^{\beta-1}$. Vi ser at hvis β er større enn 1 så vil dødsraten øke med alder. Det er dette som kalles aldring i biologien.