TAMS79: Föreläsning 9
Approximationer och stokastiska processer

Johan Thim*

18 november 2018

9.1 Binomialfördelning

Vi har redan stött på denna fördelning flera gånger. Situationen är att ett slumpförsök har två möjliga utfall, ett med sannolikhet p och det andra med $1-p$. Vi upprepar försöket oberoende n gånger, och räknar antalet X gånger det första utfallet inträffar. Vi kallar X för binomialfördelad med parametrarna n och p, och skriver $X \sim \mathrm{Bin}(n,p)$.

Binomialfördelning

Sats. Om $X \sim \mathrm{Bin}(n,p)$, så är $E(X) = np$ och $V(X) = np(1-p)$. Vidare gäller att vi kan approximera $X \overset{\text{appr.}}{\sim} \mathrm{N}(np, \sqrt{np(1-p)})$ om $np(1-p) \geq 10$.

Beviset för väntevärde och varians följer av argumentet i föregående föreläsning angående **CGS**. Vi skriver X som en summa av n oberoende Bernoulli-variabler $X_k \sim \mathrm{Be}(p)$, så väntevärdet $E(X) = \sum_{k=1}^{n} E(X_k) = np$, eftersom $E(X_k) = 0 \cdot (1-p) + 1 \cdot p = p$. På samma sätt, $V(X) = np(1-p)$ eftersom $V(X_k) = 0^2 \cdot (1-p) + 1^2 \cdot p - p^2 = p(1-p)$.
När det gäller approximationen till normalfördelning så följer detta av CGS. Att just kravet $np(1-p) \geq 10$ ger en bra approximation kräver en lite djupare analys av på vilket sätt sannolikheterna konvergerar.

Exempel

Vi sår 1000 stycken frön som har en grobarhet på 80% (sannolikheten att ett frö gror). Vad är sannolikheten att högst 180 stycken inte gror?

Låt X vara antalet frön som inte gror. Vi antar att olika frön är oberoende av varandra. Då är $X \sim \mathrm{Bin}(1000, 0.2)$. Eftersom

$$np(1-p) = 1000 \cdot 0.2 \cdot 0.8 = 160 \gg 10,$$

så är $X \overset{\text{appr.}}{\sim} \mathrm{N}(200, \sqrt{160})$. Vi beräknar

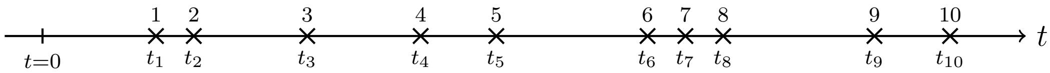
$$P(X \leq 180) \approx \Phi((180-200)/\sqrt{160}) = \Phi(-1.5811) = 1 - \Phi(1.5811) = 0.057.$$

Det är alltså ca 6% sannolikhet. Verklig sannolikhet (MATLAB `binocdf(180,1000,0.2)`) är 6.02%.

*jothi@mai.liu.se

9.2 Poissonfördelning

Antag att vi har en situation där händelser inträffar oberoende av varandra med en konstant intensitet λ, det vill säga, på t tidsenheter inträffar i genomsnitt λt händelser. Denna typ av situation brukar ofta moduleras med hjälp av *Poisson*-fördelningen. Om $X(t)$ är antalet händelser i tidsintervallet $[0, t]$, så säger vi att $X(t)$ är Poissonfördelad med väntevärde $\mu = \lambda t$.



Händelser (markerade med kryss och numrerade) i tidsintervallet $[0, t]$. Tiderna t_k är tidpunkten för händelsen k.

Poissonfördelning

Sats. Vi kallar X för Poissonfördelad med parametern μ, $X \sim \mathrm{Po}(\mu)$, om sannolikhetsfunktionen ges av

$$p_X(k) = P(X = k) = \frac{\mu^k e^{-\mu}}{k!}, \quad k = 0, 1, 2, \ldots .$$

Variabeln X har $E(X) = V(X) = \mu$ (samma väntevärde som varians, parametern μ).

Hur hänger situationen ovan ihop med definitionen av p_X? Vi fixerar tiden t och delar in intervallet $[0, t]$ i n lika stora delar, där vi väljer n så stort att det högst finns en händelse i varje delintervall. Vi introducerar en sannolikhet p som är sannolikheten att ett visst delintervall innehåller en händelse. Det är samma p för alla delintervall och sambandet $np = \lambda t$ måste gälla. Eftersom händelserna är oberoende måste $X(t) \sim \mathrm{Bin}(n, p)$. Egenskaper för binomialfördelningen medför att $E(X(t)) = np = \lambda t$.
Vi börjar med att betrakta fallet med noll händelser i intervallet $[0, t]$, det vill säga, händelsen att $X(t) = 0$:

$$P(X(t) = 0) = \binom{n}{0} p^0 (1-p)^{n-0} = \left(1 - \frac{\lambda t}{n}\right)^n = \exp\left(n \ln\left(1 - \frac{\lambda t}{n}\right)\right) \to e^{-\lambda t},$$

då $n \to \infty$. Här har vi utnyttjat standardgränsvärdet $s^{-1}\ln(1 + s) \to 1$ då $s \to 0$.
I det allmänna fallet kan vi visa att

$$P(X(t) = k) \to \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad n \to \infty.$$

För att se detta, låt k vara fix och betrakta

$$\begin{aligned} P(X(t) = k) &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{(n-k)!k!} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{(\lambda t)^k}{k!} \cdot \frac{\left(1 - \frac{\lambda t}{n}\right)^n}{\left(1 - \frac{\lambda t}{n}\right)^k} \to 1 \cdot \frac{(\lambda t)^k}{k!} \cdot \frac{e^{-\lambda t}}{1}, \quad n \to \infty. \end{aligned}$$

Vad detta innebär är att om $X_n \sim \mathrm{Bin}(n, \lambda/n)$, så kommer $X_n \xrightarrow{\mathrm{D}} X \sim \mathrm{Po}(\lambda)$. Detta följer av satsen på föregående föreläsning om att konvergens i fördelning blir ekvivalent med konvergens av sannolikhetsfunktioner i det diskreta fallet.
Detta ger oss även en användbar approximationssats för binomialfördelningen.

Approximation: Binomial till Poisson

Sats. Om $X \sim \text{Bin}(n, p)$ med $n \geq 10$ och $p \leq 0.1$, så är $X \overset{\text{appr.}}{\sim} \text{Po}(np)$.

Att p_X verkligen är en sannolikhetsfunktion följer från Maclaurinuteckling av e^x:

$$\sum_{k=0}^{\infty} p_X(k) = e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = e^{-\mu} \cdot e^{\mu} = 1.$$

Låt oss även härleda väntevärde och varians. För väntevärdet:

$$\sum_{k=0}^{\infty} k p_X(k) = e^{-\mu} \left(0 + \sum_{k=1}^{\infty} \frac{\mu^k}{(k-1)!} \right) = \mu e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} = \mu e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = \mu.$$

Variansen är lite bökigare. Vi kan inte direkt räkna ut $E(X^2)$, utan tar till omskrivningen

$$E(X^2) = E(X(X-1)) + E(X).$$

Alltså,

$$\begin{aligned} E(X(X-1)) &= \sum_{k=0}^{\infty} k(k-1) p_X(k) = e^{-\mu} \left(\sum_{k=2}^{\infty} \frac{\mu^k}{(k-2)!} \right) = \mu^2 e^{-\mu} \sum_{k=2}^{\infty} \frac{\mu^{k-2}}{(k-2)!} \\ &= \mu^2 e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = \mu^2, \end{aligned}$$

så $E(X^2) = \mu^2 + \mu$, vilket medför att $V(X) = E(X^2) - E(X)^2 = \mu$.

Exempel

Sågaren Sverker sågar ut brädor som har normalfördelad längd $L \sim \text{N}(200, \sqrt{50})$ ($\sigma^2 = 50$), enhet: cm. Om Sverker en vacker dag sågar upp 300 brädor (oberoende av varandra), vad är sannolikheten att färre än 5 stycken är kortare än 185 cm?

Lösning: Vi räknar först ut sannolikheten p att en bräda är kortare än 185 cm:

$$p = P(L < 185) = P\left(\frac{L - 200}{\sqrt{50}} < \frac{-15}{\sqrt{50}} \right) = \Phi(-2.12) = 1 - \Phi(2.12) = 0.0170.$$

Låt X vara antalet brädor av 300 som är kortare än 185 cm. Det följer att $X \sim \text{Bin}(300, p)$. Sannolikheten p är alltså liten, och $300p(1-p) = 5.01$ är betydligt mindre än 10, så normalapproximation fungerar antagligen inte. Men Poissonapproximation borde fungera bra då $p \ll 0.1$ och $n = 300 \gg 10$. Alltså är $X \overset{\text{appr.}}{\sim} \text{Po}(300 \cdot 0.0170) = \text{Po}(5.10)$. Ur tabell (interpolation mellan Po(5.0) och Po(5.2)):

$$P(X \leq 4) \approx \frac{0.4405 + 0.4061}{2} = 0.4233.$$

Alltså ungefär 42% chans. Verklig sannolikhet blir 42.15%. Normalapproximation skulle i fallet ge 31%, vilket är alldeles för lågt.

Addition av oberoende Poissonfördelade variabler

Sats. Låt $X \sim \text{Po}(\mu_1)$ och $Y \sim \text{Po}(\mu_2)$ vara oberoende. Då är $X + Y \sim \text{Po}(\mu_1 + \mu_2)$.

Satsen förefaller intuitivt att vara rimlig. Vi lägger helt enkelt ihop händelserna från två liknande processer, det förväntade antalet blir nu $\mu_1 + \mu_2$, och på grund av beteendet hos var och en tippar vi att summan fungerar på samma sätt. Formellt kan vi visa satsen medelst den så kallade faltningssatsen. Den simultana sannolikhetsfunktionen för (X, Y) ges av produkten $p_X(i)p_Y(j)$, och vi söker sannolikhetsfunktionen för $Z = X + Y$. Alltså,

$$p_Z(k) = P(X + Y = k) = \sum_{i+j=k}\sum p_X(i)p_Y(j) = \sum_{i=0}^{k} p_X(i)p_Y(k-i).$$

Dubbelsumman blir en enkelsumma eftersom vi bara summerar över "diagonalen" (när k är fixt och $i + j = k$). Sen är $p_X(i) = 0$ då $i < 0$ och $p_Y(k - i) = 0$ då $i > k$ så det räcker att summera från $i = 0$ till $i = k$. Vidare,

$$\begin{aligned} p_Z(k) &= \sum_{i=0}^{k} e^{-\mu_1}\frac{\mu_1^i}{i!}e^{-\mu_2}\frac{\mu_2^{k-i}}{(k-i)!} = \frac{e^{-(\mu_1+\mu_2)}}{k!}\sum_{i=0}^{k}\frac{k!}{(k-i)!i!}\mu_1^i\mu_2^{k-i} \\ &= \frac{e^{-(\mu_1+\mu_2)}}{k!}\sum_{i=0}^{k}\binom{k}{i}\mu_1^i\mu_2^{k-i} = \frac{e^{-(\mu_1+\mu_2)}}{k!}(\mu_1+\mu_2)^k, \end{aligned}$$

där vi utnyttjat binomialsatsen i sista steget. Detta uttryck är inget annat än sannolikhetsfunktionen för en $\text{Po}(\mu_1 + \mu_2)$-fördelad variabel, vilket var precis det vi ville visa!

Denna sats kan vi använda för att dela upp en $\text{Po}(\mu)$ fördelad variabel i $\lfloor \mu \rfloor$ stycken oberoende variabler med väntevärde ett, och en liten svans (med längd $\mu - \lfloor \mu \rfloor$). På detta sätt kan man visa följande sats.

Approximation av Poissonfördelning

Sats. Låt $X \sim \text{Po}(\mu)$ med $\mu \geq 15$. Då är $X \overset{\text{appr.}}{\sim} \text{N}(\mu, \sqrt{\mu})$ $(V(X) = \mu)$.

Exempel

Låt X vara antal paket i en datakö under en sekund. Mätningar har visat att en vettig modell är $X \sim \text{Po}(250)$. Beräkna approximativt $P(X < 240)$.

Lösning: Eftersom väntevärdet $\mu = 250 \gg 15$ så kan vi normalapproximera. Då blir

$$P(X < 240) = P(X \leq 239) \approx \Phi\left(\frac{239 - 250}{\sqrt{250}}\right) = \Phi(-0.6957) = 1 - \Phi(0.6957) = 0.2433.$$

Exakt värde: 0.2552.

9.3 Exponentialfördelning

Vi fortsätter med situationen för Poissonfördelningen, men istället för att räkna antalet händelser räknar vi tiden mellan händelser. Det visar sig nämligen att tiden är *exponentialfördelad.* Varför? Betrakta följande. Låt T_1 vara tiden till den första händelsen. Detta innebär att inga händelser inträffar på tiden t, dvs $P(T_1 > t) = P(X(t) = 0)$. Men $P(X(t) = 0) = e^{-\lambda t}$. Vi betraktar fördelningsfunktionen för T_1:

$$F_{T_1}(t) = P(T_1 \leq t) = 1 - P(T_1 > t) = 1 - e^{-\lambda t}, \quad t \geq 0.$$

Vi kan nu derivera fram täthetsfunktionen: $f_{T_1}(t) = F'_{T_1}(t) = \lambda e^{-\lambda t}$ för $t \geq 0$. Detta är precis täthetsfunktionen för en $\mathrm{Exp}(\lambda)$-fördelad variabel!

Exponentialfördelning

Sats. En stokastisk variabel X med täthetsfunktion $f_X(x) = \lambda e^{-\lambda x}$, $x \geq 0$, kallas exponentialfördelad. Väntevärde och varians ges av $E(X) = \lambda^{-1}$ och $V(X) = \lambda^{-2}$.

En intressant egenskap hos exponentialfördelningen är att den är minnesslös:

$$\begin{aligned} P\big(X > x + x_0 \mid X > x_0\big) &= \frac{P(\{X > x + x_0\} \cap \{X > x_0\})}{P(X > x_0)} = \frac{P(X > x + x_0)}{P(X > x_0)} = \frac{e^{-\lambda(x_0+x)}}{e^{-\lambda x_0}} \\ &= e^{-\lambda x} = P(X > x). \end{aligned}$$

9.4 Stokastiska processer

Stokastisk process

Definition. En **stokastisk process** är en familj $\{X_t\}_{t\in I}$ av stokastiska variabler $X_t \colon \Omega \to E$, där t är ett index i indexmängden I och processen tar värden i **tillståndsrummet** E. Ibland skriver vi $\{X(t)\}_{t\in I}$ istället om det inte finns risk för missförstånd.

Följande kategorisering kan göras. Processen $\{X_t\}_{t\in I}$ säges

(i) ha **diskret tid** om I är uppräknelig, typiskt $I = \{0, 1, 2, 3, \ldots\}$;

(ii) ha **kontinuerlig tid** om I är ett kontinuum, typiskt $I \subset [0, \infty)$ ett intervall;

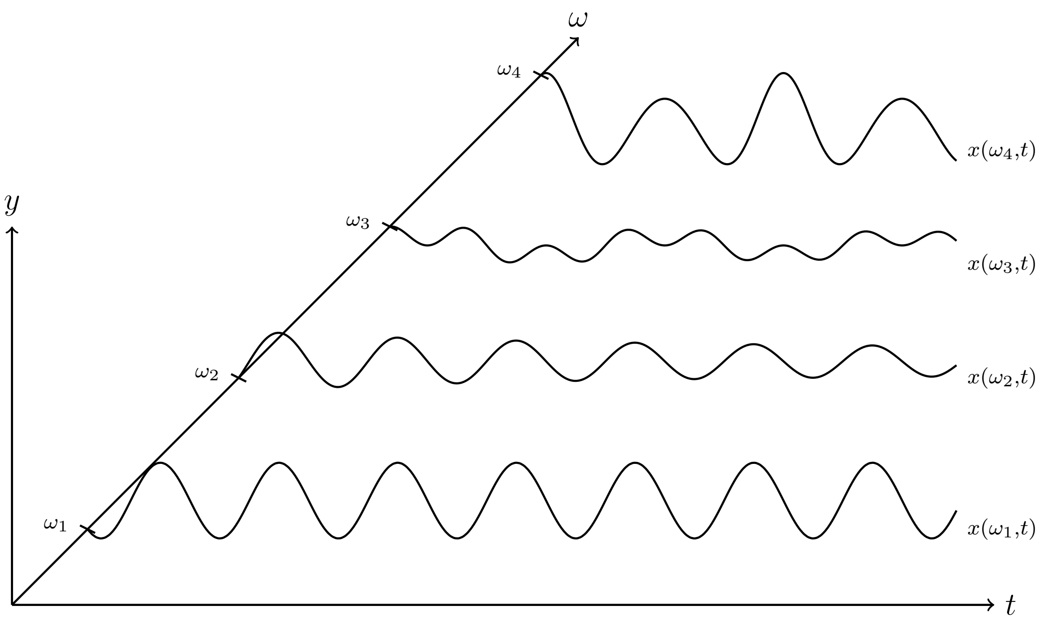
(iii) vara **diskret** om tillståndsrummet E är uppräkneligt: $E = \{E_0, E_1, E_2, \ldots\}$;

(iv) vara **kontinuerlig** om tillståndsrummet E är ett kontinuum, typiskt $E \subset [0, \infty)$ igen.

Processer kan alltså vara både kontinerliga eller diskreta, och ha kontinuerlig eller diskret tid. Speciellt kan vi till exempel ha en diskret process med kontinuerlig tid (tillståndsrummet uppräkningsbart men tiden ett kontinuum).

Anmärkning: eftersom X_t är en stokastisk variabel så borde vi kräva att $E \subset \mathbf{R}$ eftersom vi definierat termen stokastisk variabel på det sättet. Vi kommer i de närmaste avsnitten att tillåta tillståndsrummet att vara mer abstrakt, säg $E = \{\text{solen skiner}, \text{det är natt}\}$ till exempel. Detta underlättar i våra tilltänkta tillämpningar.

För varje $\omega \in \Omega$ kan vi prata om **realiseringen** av processen i form av en tidsfunktion $x(\omega, t)$ som avbildar indexmängden I in i tillståndsrummet E; $x(\omega, \cdot\,) \colon I \to E$. Med andra ord, $x(\omega, t)$ ger värdet av utfallet vid tiden t.



Några realiseringar vid fixerade utfall ω_i för en tidskontinuerlig kontinuerlig stokstisk process.

Exempel

(i) $X_t = s_t + N_t$ där $N_t \sim (0, \sigma)$ för alla t är oberoende. Detta är en signal s_t vi ofta är intresserade av som är störd av brus (som är normalfördelat i detta fall).

(ii) $X_t = A\cos 2t + Be^{-t}$ där A och B är stokastiska variabler. Vad händer när t går mot ∞?

Det finns mycket teori om stokastiska processer; hur man klassificerar dem (stationär? svagstationär? etc) och hur man studerar beteendet. Detta ligger utanför ramen på denna kurs, men vi skall punktstudera vissa specialfall den närmaste tiden. Först ut är den så kallade Poissonprocessen.

9.5 Poissonprocessen

Vi återkommer nu till poissonfördelning, men på ett lite annorlunda sätt. Låt $X(t)$ räkna antalet impulser i intervallet $[0, t]$. Vi ställer följande krav.

Poissonprocess

Definition. En stokastisk process $\{X(t)\}_{t\in[0,\infty)}$ är en **Poissonprocess** med intensitet λ om och endast om

(i) Antalet impulser i disjunkta tidsintervall är oberoende stokastiska variabler;

(ii) $P(X(t+h) - X(t) = 1) = \lambda h + o(h)$;

(iii) $P(X(t+h) - X(t) \geq 2) = o(h)$.

Notationen $o(h)$ betyder något som går mot noll snabbare än h. Om $f(h) = o(h)$ innebär det att $\lim_{h\to 0} \frac{f(h)}{h} = 0$. Kommer ni ihåg stora *ordo* från resttermer i Maclaurinutvecklingar i analysen? Vi hanterar lilla *ordo* på liknande sätt och kan göra analoga "förenklingar."
Så hur hänger den här definitionen ihop med Poissonfördelningen?
Låt $p_n(t) = P(X(t) = n)$ och anta att $h > 0$ är litet. Eftersom $[0, t)$ och $[t, t+h)$ är disjunkta tidsintervall så är variablerna $X(t)$ och $X(t+h) - X(t)$ oberoende, och $X(t+h)$ är alltså summan av dessa två oberoende variabler. Händelsen att $X(t+h) = n$ kan nu delas upp i unionen

$$\{X(t+h) = n\} = \cup_{k=0}^{n} \big(\{X(t) = n-k\} \cap \{X(t+h) - X(t) = k\}\big),$$

där händelserna i varje snitt är oberoende! Vidare medför våra villkor att

$$P(X(t+h) - X(t) = 0) = 1 - \lambda h + o(h).$$

Alltså blir

$$\begin{aligned}
p_n(t+h) &= P(X(t) = n)P(X(t+h) - X(t) = 0) \\
&\quad + P(X(t) = n-1)P(X(t+h) - X(t) = 1) \\
&\quad + \sum_{k=2}^{n} P(X(t) = n-k)P(X(t+h) - X(t) = k) \\
&= p_n(t)(1 - \lambda h + o(h)) + p_{n-1}(t)(\lambda h + o(h)) + \sum_{k=2}^{n} p_{n-k}(t)o(h) \\
&= p_n(t)(1 - \lambda h + o(h)) + p_{n-1}(t)(\lambda h + o(h)) + o(h).
\end{aligned}$$

Vi stuvar om lite och finner att

$$\frac{p_n(t+h) - p_n(t)}{h} = -\lambda p_n(t) + \lambda p_{n-1}(t) + \frac{o(h)}{h}.$$

Vi låter $t \to 0^+$ och erhåller då att högerderivatan uppfyller

$$(p_n)'_+(t) = -\lambda p_n(t) + \lambda p_{n-1}(t), \quad n = 1, 2, \ldots.$$

Speciellt gäller $p_0'(t) = -\lambda p_0(t)$, $t \geq 0$, för $n = 0$. Med det naturliga begynnelsevillkoret $p_0(0) = 1$ har denna differentialekvation lösningen $p_0(t) = e^{-\lambda t}$. Om vi löser differentialekvationen för $n \geq 1$ erhåller vi

$$p_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda}, \quad n = 0, 1, 2, 3, \ldots.$$

Detta är sannolikhetsfunktionen för Poissonfördelningen! Vi har med andra ord precis visat att $X(t) \sim \mathrm{Po}(\lambda t)$ för alla $t > 0$.

Från detta och tidigare härledda resultat för Poissonfördelningen har vi nu följande resultat för Poissonprocessen.

Egenskaper hos Poissonprocessen

Låt $X(t)$ vara en Poissonprocess med intensitet $\lambda > 0$.

(i) Vid tiden $t > 0$ gäller att $X(t) \sim \text{Po}(\lambda t)$.

(ii) Tiden T mellan två händelser är exponentialfördelad: $T \sim \text{Exp}(\lambda)$.

(iii) Tider mellan olika händelser är oberoende stokastiska variabler.

(iv) Sammanslagningen av två oberoende Poissonprocesser med intensiteter λ_1 respektive λ_2 är en Poissonprocess med intensitet $\lambda_1 + \lambda_2$.

(v) Om $p_1, p_2 \in [0, 1]$ med $p_1 + p_2 = 1$, så kan $X(t)$ delas upp i två oberoende Poissonprocesser med intensiteter $p_1\lambda$ respektive $p_2\lambda$.