Kapitel 7

Konvergens och Kontinuitet

Gränsvärdesbegreppet är väldigt centralt inom matematik. Som du förhoppningsvis kommer ihåg från matematisk analys så definieras tex derivatan av en funktion $f : \mathbb{R} \mapsto \mathbb{R}$ i punkten $x_0 \in \mathbb{R}$ som gränsvärdet

$$\frac{f(x_0 + h) - f(x_0)}{h} \quad \text{då } h \to 0$$

och om det existerar, kallar vi det för derivatan i punkten $x_0 \in \mathbb{R}$ av funktionen f och betecknar den $f'(x_0)$.

Vad menas då med limes-utsagan ovan? Jo, att för varje $\epsilon > 0$ så finns det ett $\delta > 0$ så att $|\frac{f(x_0+h)-f(x_0)}{h} - f'(x_0)| < \epsilon$ för $|h| < \delta$. De två termerna innanför absolutbeloppet kan alltså komma godtyckligt nära varandra genom att välja ett tillräckligt litet h

Matematiker faller i två kategorier. De som älskar konvergens, och de som inte älskar konvergens. De förra kallas för analytiker, och de senare för algebraiker. Sannolikhetsteorin är en del av den matematiska analysen, alltså så är konvergens det bästa vi sannolikhetsteoriker vet.

Kärt barn har ju som bekant många namn, och redan från den grundläggande analysen har ni säkert hört talas om vanlig konvergens, Cauchy konvergens, absolut konvergens, punktvis konvergens, likformig konvergens, etc. etc. ad infinum. Det visar sig dock att dessa inte räcker för stokastiska processer, utan att man behöver inte mindre än fyra olika sorters konvergens till.

Låt $\xi_1, \xi_2, \xi_3, \ldots$ vara en sekvens av stokastiska variabler (med andra ord en tidsdiskret stokastisk process). Redan i grundkursen så gör man gränsvärdes observationer om dessa, till exempel stora talens lag som säger att om alla ξ_k är oberoende och har samma fördelning, så "går"

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

mot $\mathbf{E}[\xi_k]$. Men vad betyder detta "går mot" egentligen?

Vi vet att ξ_k egentligen är en funktion $\xi_k(\omega)$, och om vi drar oss till minnes från analysen, så säger man att en sekvens funktioner $f_1, f_2, \ldots$ konvergerar punktvis mot f om $\lim_{k\to\infty} f_k(x) = f(x)$ för alla x. Är punktvis konvergens det vi letar efter?

Tänk nu tillbaka till det sannolikhetsrum som vi konstruerade i detalj i en tidigare föreläsning, där utfall ω är oändliga binära strängar av typen $010100111001\ldots$, Ω är mängden av alla sådana strängar och sannolikhetsmåttet tilldelar sannolikheten p att varje koordinat blir 1, oberoende av alla andra. Låt $\xi_k(\omega) = \omega_k$ (den k-te binära koordinaten). Således är ξ_k oberoende och likafördelad för alla k, och $\mathbf{E}[\xi_k] = p$.

Gäller nu att $\frac{1}{n}\sum_{k=1}^{n}\xi_k(\omega) \to p$ för alla $\omega \in \Omega$? Självklart inte! $111111\ldots$ (strängen med bara ettor) tillhör ju Ω, men för den blir genomsnittet alltid 1 även om p är litet. Alltså är punktvis konvergens ett alldeles för starkt krav för att satser som stora talens lag ska kunna gälla i allmänhet.

Konvergensidén vi behöver måste vara grundläggande probabilistiskt. Det är inte så att det inte finns några ω för vilka konvergens inte gäller: det viktiga är att sannolikheten för att vi ska få ett sådant ω är 0. (Slug student kanske nu tycker: varför plockar vi inte bara bort sådana "dåliga" ω ur utfallsmängden. Slug lärare svarar: Hurdå?)

Nästan säker konvergens

Denna typ av konvergens motiveras av exemplet ovan. Låt $\xi_1, \xi_2, \ldots$ vara en sekvens av stokastiska variabler, och ξ en enda sådan. Då säger vi att $\xi_k \xrightarrow{n.s.} \xi$ *nästan säkert (n.s.)* då $k \to \infty$ om:

$$\mathbf{P}(\xi_k \to \xi) = \mathbf{P}(\{\omega \in \Omega : \xi_k(\omega) \to \xi(\omega)\}) = 1.$$

Stora talens lag är ett exempel på sådan konvergens: satsen att stora talens lag gäller n.s. kallas "stora talens starka lag" ("strong law of large numbers").

Observera att "nästan säker" egentligen är ett felaktigt namn. Man använder "nästan" för att påminna om att det är en svagare form av konvergens än punktvis, men faktum är att det är "säkert" och inte "nästan säkert" att sekvensen konvergerar!

Vi kommer inte att använda nästan säker konvergens mycket i den här kursen, men det tillhör en allmän kunskap i sannolikhetsteori att förstå begreppet.

Konvergens i sannolikhet

Man säger att $\xi_k \xrightarrow{\mathbf{P}} \xi$ *i sannolikhet* om $\mathbf{P}(|\xi_k - \xi| > \epsilon) \to 0$ för alla ϵ. Konvergens i sannolikhet säger alltså att sannolikheten att ξ_k och ξ skiljer sig "mycket" går mot 0 när k går mot oändligheten. Det är en svagare form konvergens en nästan säker, och oftast lättare att visa: statsen att stora talens lag gäller i sannolikhet kallas stora talens svaga lag ("weak law of large numbers").

Vi kommer inte att använda sannolikhet i konvergens mycket (eller äns alls) i den här kursen.

Konvergens i kvadratisk medel

Vi säger att $\xi_k \xrightarrow{L^2} \xi$ *i kvadratisk medel* om $\mathbf{E}\{(\xi_k - \xi)^2\} \to 0$. Det är ganska lätt att se att om en sekvens stokastiska variabler konvergerar i kvadratiskt medel, så gör de även det i sannolikhet.

Konvergens i kvadratisk medel är något som används ganska ofta i teorin om stokastiska processer, och det är den enda typ av konvergens som används utförligt i kursboken. Mer om den nedan.

Konvergens i Fördelning

Den sista typen av konvergens, skriven $\xi_k \xrightarrow{D} \xi$ är *konvergens i fördelning*, eller svag konvergens, som det också kallas. Sådan konvergens gäller om $F_{\xi_k} \to F_\xi$ punktvis. Observera att konvergens i fördelning inte egentligen säger vad själva gränsvärdet av ξ_k är, bara vad det har för fördelning. Eftersom vi vet två stokastisk variabler kan ha samma fördelning utan att någonsin vara lika, så kan det gälla att $\xi_k \xrightarrow{D} \xi$ och att $\xi_k \xrightarrow{D} \eta$ även om $\mathbf{P}(\xi \neq \eta) = 1$. Inte konstigt att denna typ av konvergens kallas svag!

Eftersom denna typ av konvergens handlar om fördelningar, så struntar man ofta i att konstruera ξ och skrivar bara t.ex. $\xi_k \xrightarrow{D} \exp(\lambda)$ för att visa att konvergensen är mot en exponentialfördelning.

Det utan tvekan viktigaste fallet av konvergens i fördelning är centrala gränsvärdes satsen, som säger att:

$$\frac{\sum_{k=1}^{n}(\xi_k - \mu)}{\sqrt{n\sigma^2}} \xrightarrow{D} N(0,1)$$

när $n \to \infty$ om ξ_k är en sekvens med oberoende, likafördelade s.v. med väntevärde μ och varians σ^2. Vi kan inte säga vilken stokastiska variabel genomsnittet konvergerar mot, men vi vet vad fördelningen närmar sig.

Att denna typ av konvergens är svag jämnfört med de andra betyder inte att den är oanvändbar: snarare tvärt om, mycket av sannolikhetsteorin och nästan hela statistiken bygger på teorin om svag konvergens. Vi kommer att använda konvergens i fördelning i den formella motiveringen för Wienerprocessen, och sedan även när vi diskuterar Markovkedjor.

Vi kommer inte att hänga upp oss så mycket på konvergens i den här kursen. Om någon tycker att det verkar spännande (och det är det!) så finns det hur mycket som helst att lära sig om varje typ: på högre nivå finns hela kurser i varje typ av konvergens.

Vidare om konvergens i kvadratiskt medel

Def. 2.6 För $\mathbb{R}$-värda s.v. $\xi, \xi_1, \xi_2, \ldots$ säger vi att ξ_k konvergerar i kvadratiskt medel mot ξ då $k \to \infty$, och skriver $\xi_k \xrightarrow{L^2} \xi$ eller $\underset{k\to\infty}{\text{l.i.m.}}\, \xi_k = \xi$ (l.i.m.="limit in mean"), om det gäller att

$$\mathbf{E}\{\xi^2\} < \infty$$

samt att

$$\mathbf{E}\{(\xi_k - \xi)^2\} \to 0 \quad \text{då } k \to \infty$$

Sats 2.3 (Momentsatsen) Antag att $\underset{k\to\infty}{\text{l.i.m.}}\, \xi_k = \xi$. Då har vi

$$\begin{aligned} \mathbf{E}\{\xi_k\} &\to \mathbf{E}\{\xi\} \\ \mathbf{Var}\{\xi_k\} &\to \mathbf{Var}\{\xi\} \\ \mathbf{E}\{\xi_k^2\} &\to \mathbf{E}\{\xi^2\} \end{aligned}$$

Om dessutom $\underset{k\to\infty}{\text{l.i.m.}}\, \eta_k = \eta$, så har vi också

$$\begin{aligned} \mathbf{E}\{\xi_k\eta_k\} &\to \mathbf{E}\{\xi\eta\} \\ \mathbf{Cov}\{\xi_k, \eta_k\} &\to \mathbf{Cov}\{\xi, \eta\} \end{aligned}$$

Ett användbart sätt att avgöra om en sekvens av stokastiska variabler konvergerar (i kvadratisk mening) mot någon stokastisk variabel får vi nedan genom Cauchy- och Loéve-kriteriet.

Från analysen kommer du kanske ihåg att en sekvens av reella tal $x_1, x_2, \dots$ är en Cauchy-sekvens om $\lim\limits_{k,l \to \infty} |x_k - x_l| = 0$. Följande sats säger att alla Cauchysekvenser av kvadritiskt integrerbarar[1] stokastiska variabler ξ_k konvergerar i kvadratiskt medel mot en kvadratiskt integrerbar s.v. ξ.

Sats 2.4 För s.v. $\xi_1, \xi_2, \dots$ med $\mathbf{E}\{\xi_k^2\} < \infty$ för $k \in \mathbb{N}$, gäller att $\underset{k\to\infty}{\text{l.i.m.}}\, \xi_k = \xi$ för någon s.v. ξ om och endast om

$$\lim_{k,l \to \infty} \mathbf{E}\{(\xi_k - \xi_l)^2\} = 0$$

Ett ekvivalent kriterium är Loévekriteriet

$$\boxed{\underset{k\to\infty}{\text{l.i.m.}}\, \xi_k \text{ existerar} \quad \Longleftrightarrow \quad \lim_{k,l\to\infty} \mathbf{E}\{\xi_k \xi_l\} \text{ existerar}}$$

som dessutom ger följande:

Sats 2.5 För s.v. $\xi_1, \xi_2, \dots$ med $\mathbf{E}\{\xi_k^2\} < \infty$ gäller att

$$\sum_{k=1}^{\infty} \xi_k \text{ konvergent} \quad \Longleftrightarrow \quad \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \mathbf{E}\{\xi_k \xi_l\} < \infty$$

Vidare har vi att

$$\sum_{k=1}^{\infty} \xi_k \text{ konvergent} \quad \Longrightarrow \quad \mathbf{E}\{\sum_{k=1}^{\infty} \xi_k\} = \sum_{k=1}^{\infty} \mathbf{E}\{\xi_k\}$$

Om dessutom $\eta_1, \eta_2, \dots$ är s.v. med $\mathbf{E}\{\eta_k^2\} < \infty$ gäller också att

$$\sum_{k=1}^{\infty} \xi_k \text{ och } \sum_{k=1}^{\infty} \eta_k \text{ konvergenta} \quad \Longrightarrow \quad \mathbf{Cov}\{\sum_{k=1}^{\infty} \xi_k, \sum_{k=1}^{\infty} \eta_k\} = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \mathbf{Cov}\{\xi_k, \eta_l\}$$

Notera att vi med $\sum_{k=1}^{\infty} \xi_k$ konvergent menar att $\sum_{k=1}^{n} \xi_k = Y_n \xrightarrow{L^2} Y_\infty = \sum_{k=1}^{\infty} \xi_k$ då $n \to \infty$.

Kontinuitet i Kvadrisk Medel

Eftersom vi nu har ett konvergenssätt för stokastiska variabler så kan vi genast definiera kontinuitet av en stokastisk process. Vi håller oss till processer där tidsparameterrummet, T, är kontinuerligt så att kontinuitet för processen har en naturlig mening.

Def 3.2 En stokastisk process $\{X(t)\}_{t \in T}$ är kontinuerlig i $t_0 \in T$ om $X(t) \xrightarrow{L^2} X(t_0)$ då $t \to t_0$, dvs om $\mathbf{E}\{X(t_0)^2\} < \infty$ samt $\mathbf{E}\{(X(t) - X(t_0))^2\} \to 0$ då $t \to t_0$.

Nu kommer ett exempel som till en början antagligen verkar förvirrande:

[1] Att en stokastisk variabel ξ är kvadratiskt integrerbar betyder att $\mathbf{E}\{\xi^2\} < \infty$.

Exempel Låt $\{X(t)\}_{t\geq 0}$ vara en Poissonprocess med intensitet λ och tag godtyckligt $t_0 \geq 0$. Vi har

$$\mathbf{E}\{X(t_0)^2\} = \mathbf{Var}\{X(t_0)\} + \big(\mathbf{E}\{X(t_0)\}\big)^2 = \lambda t_0 + (\lambda t_0)^2 < \infty$$

samt

$$\begin{aligned}
\mathbf{E}\{\big(X(t) - X(t_0)\big)^2\} &= \mathbf{Var}\{(X(t) - X(t_0)\} + \big(\mathbf{E}\{X(t) - X(t_0)\}\big)^2 = \\
&= V_X(t) + V_X(t_0) - 2r_X(t, t_0) + (\lambda(t - t_0))^2 = \\
&= \lambda(t + t_0 - 2\min(t, t_0)) + \lambda^2(t - t_0)^2 = \\
&= \lambda|t - t_0| + \lambda^2(t - t_0)^2 \;\; \to \;\; 0 \quad \text{då} \;\; t \to t_0
\end{aligned}$$

Poissonprocessen är alltså kontinuerlig i kvadratisk mening.

Men hur kan detta stämma? Varje realisering $X(\omega_0, t)$ av en Poissonprocess är ju en stegvis ökande funktion, och sådanna är inte kontinuerliga.

Svaret är att vår definition av kontinuitet faktiskt inte kräver att realiseringarna skall vara kontinuerliga. Vår definition går ju ut på att vi först tar väntevärdet av kvadratskillnaden, och *sedan* låter tidsparametern närma sig, på så sätt medelvärdesbildar vi bort hoppen som man ser i varje enskild realisering.