1 Sannolikheter

Terminologi: Ett **experiment** leder till ett **utfall**. Mängden av alla möjliga utfall kallas för **utfallsrummet** och betecknas med S. En **händelse** A består av ett eller flera utfall och är en delmängd av S.

Ett sannolikhetsmått P är en funktion som är definierad på händelser och uppfyller **sannolikhetsaxiomen**

1. $P(S) = 1$ (något av utfallen inträffar säkert).
2. $P(A) \geq 0$ för alla händelser A.
3. $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ om $A_1, A_2, \ldots$ är **disjunkta**, dvs inte kan inträffa samtidigt.

Satser som följer av axiomen:

1. $P(A^C) = 1 - P(A)$.
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Den **betingade** sannolikheten för A givet att B har inträffat ges av $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

$B_1, \ldots, B_n$ är en **partition** av S om $\cup_{i=1}^{n} B_i = S$ och alla $B_1, \ldots, B_n$ är disjunkta, dvs precis en av händelserna B_i inträffar. Ex: $\{B, B^C\}$ är en partition.

Lagen om **total sannolikhet** säger att
$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$ då $B_1, \ldots, B_n$ är en partition.

Bayes sats:
$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$.

Två händelser A och B är **oberoende** om $P(A \cap B) = P(A)P(B)$.

2 Stokastiska variabler

En **stokastiska variabel** (sv) X är en variabel vars värde är slumpmässigt. En sv är **diskret** om den endast kan anta ett uppräkneligt antal värden (ex: bara heltal), och **kontinuerlig** om den kan anta ett överuppräkneligt antal värden (ex: alla reella talen) men inget enskilt värde har en positiv sannolikhet.

X:s **fördelningsfunktion** ges av $F(x) = P(X \leq x)$. Om X är diskret har den en sannolikhetsfunktion $f(x) = P(X = x)$ så att $F(x) = \sum_{k \leq x} f(k)$. Om den är kontinuerlig gäller istället $F(x) = \int_{-\infty}^{x} f(u)du$, där f är en **täthetsfunktion**.

Väntevärdet för X ges av $E[X] = \sum_k kf(k)$ om X är diskret, $E[X] = \int_{-\infty}^{\infty} uf(u)du$ om X är kontinuerlig. Räkneregler (a konstant):

- $E[a] = a$.
- $E[aX] = aE[X]$.
- $E[X + Y] = E[X] + E[Y]$.

Markovs olikhet: $P(X \geq c) \leq \frac{E[X]}{c}$.

Man kan beräkna $E[h(X)]$ om h är någon funktion: $E[h(X)] = \sum_k h(k)f(k)$ eller $\int_{-\infty}^{\infty} h(u)f(u)du$.

Variansen ges av $\mathrm{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$.

- $\mathrm{Var}(a) = 0$.
- $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$.
- $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ om X och Y är oberoende.

Chebyshevs olikhet: $P(|X - E[X]| \geq c) \leq \frac{\mathrm{Var}(X)}{c^2}$.

Standardavvikelsen $\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$.

Momentgenererande funktionen $m_X(t) = E[e^{Xt}] = \sum_{k=0}^{\infty} E[X^k]\frac{t^k}{k!}$. $E[X^k] = m_X^{(k)}(0)$.

Två sv X och Y har en **simultan** fördelning med (om de är diskreta) simultan sannolikhetsfunktion $f_{X,Y}(x, y) = P(X = x, Y = y)$. Man kan få fram den **marginella** sannolikhetsfunktionen genom att summera bort den andra variabeln: $f_X(x) = \sum_k f_{X,Y}(x, k)$, $f_Y(y) = \sum_k f_{X,Y}(k, y)$. Väntevärden beräknas: $E[h(X, Y)] = \sum_k \sum_l h(k, l)f_{X,Y}(k, l)$. För kont. sv byts summorna mot integraler.

X och Y är **oberoende** om $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ för alla x, y. **Kovariansen** $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$ och **korrelationen** $\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X,Y)}{\mathrm{SD}(X)\mathrm{SD}(Y)}$ mäter linjärt beroende mellan X och Y. Ober. $\Rightarrow$ Cov = 0, men inte omvänt.

Om $X_1, X_2, \ldots$ alla är oberoende med samma fördelning och $\mu = E[X_1] < \infty$ så säger **Stora Talens Lag** att $\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu$ med sannolikhet 1. Den svaga stora talens lag: "$P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| > \epsilon) \to 0$ för alla $\epsilon > 0$", kan man visa med Chebyshevs olikhet.

2.1 Några fördelningar

Om alla värden i ett intervall I är lika sannolika så är X:s fördelning **likformig**. Om X är diskret: $f(k) = \frac{1}{n}, k \in I$ där n är antalet punkter i I. Om X är kont.: $f(x) = \frac{1}{L}, x \in I$ där L är I:s längd.

Om man upprepar försök som lyckas med sannolikhet p oberoende av varandra så är antalet försök till och med man lyckas första gången **geometriskt fördelat**: $f(k) = (1-p)^{k-1}p, k = 1, 2, \ldots$, och antalet lyckade försök på de n första gångerna **binomialfördelat**: $f(k) = \binom{n}{k}p^k(1-p)^{n-k}$.

Om (tid)punkter är slumpmässigt fördelade så att

1. antal punkter i disjunkta intervall är oberoende,
2. förväntat antal punkter i ett intervall är proportionellt mot intervallets längd L: λL,
3. sannolikheten för fler än en punkt i ett intervall går snabbt mot noll då intervallets längd minskar,

då är är antalet punkter i ett intervall av längd L **Poissonfördelat**: $f(k) = e^{-\lambda L}\frac{(\lambda L)^k}{k!}$. Avståndet mellan två punkter (om intervallen är längs en (tids)axel) har en **exponentialfördelning**: $f(x) = \lambda Le^{-\lambda Lx}$ (en kont. fördelning).

En kont. sv X med täthet: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ sägs vara **normalfördelad** med $E[X] = \mu, \mathrm{Var}(X) = \sigma^2$. Räkneregler för normalfördelade sv:

- $aX + b \sim N(a\mu + b, a^2\sigma^2)$
- $X+Y \sim N(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$ om X och Y är oberoende och normalfördelade.

Enligt **Centrala Gränsvärdessatsen** är en summa av många oberoende stokastiska variabler som alla har samma fördelning approximativt normalfördelad då antalet termer är stort (≥ 30 räcker ofta bra). Normalfördelningens väntevärde är summan av de enskilda väntevärdena och dess varians är summan av de enskilda varianserna.

För binomialfördelningen leder CGS till att $\mathrm{Bin}(n, p) \approx N(np, np(1-p))$ då $n \cdot \min\{p, 1-p\} \geq 5$.

3 Markovkedjor

En **stokastisk process** är en följd av sv $X_1, X_2, \ldots$ som kan ha något beroende. Om $P(X_{k+1} = j | X_k = i, X_{k-1} = i_2, \ldots) = P(X_{k+1} = j | X_k = i) = p_{ij}$, dvs "det räcker med att veta dagens tillstånd för att veta fördelningen för värdet imorgon" (vilket kallas Markovegenskapen), då är $X_1, X_2, \ldots$ en **Markovkedja**.

Man brukar samla **övergångssannolikheterna** p_{ij} i en **övergångsmatris** $P = (p_{ij})$. Elementen $p_{ij}^{(k)}$ i matrisen $P^k = P \cdots P$ ger sannolikheterna att ta sig från i till j efter k steg.

Ett tillstånd a kallas **absorberande** om $p_{aa} = 1$, dvs man fastnar i a om man kommer dit. m_i är tiden till absorbtion i något tillstånd då man startar i tillstånd i och $q_i^{(a)}$ är sannolikheten att absorberas i a, istället för någon annanstans, då man startar i tillstånd i:

$$m_i = \begin{cases} 0 & \text{om } i \text{ absorberande} \\ 1 + \sum_k p_{ik} m_k & \text{annars} \end{cases}$$

$$q_i^{(a)} = \begin{cases} 1 & \text{om } i = a \\ 0 & \text{om } i \neq a \text{ och absorberande} \\ \sum_k p_{ik} q_k^{(a)} & \text{annars} \end{cases}$$

4 Statistik

Ett **stickprov** $X_1, \ldots, X_n$ består av oberoende och likafördelade sv. En **skattare** $\hat{\theta}$ av en parameter θ är en funktion av stickprovet. En skattare bör gärna vara **väntevärdesriktig**: $E[\hat{\theta}] = \theta$, och ha låg varians. Ett utfall av en skattare kallas för en **skattning**. Vanliga väntevärdesriktiga skattare: $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ (skattar väntevärdet), $S^2 = \frac{1}{(n-1)}\sum_{i=1}^n (X_i - \bar{X})^2$ (skattar variansen), $\hat{p} = \frac{1}{n}X$, där $X \sim \mathrm{Bin}(n, p)$ (skattar p).

Ett konfidensintervall (KI) med konfidensgrad $(1-\alpha)$ är ett stokastiskt intervall som med sannolikhet $(1-\alpha)$ täcker det sanna parametervärdet θ. Många konfidensintervall är på formen $\hat{\theta} \pm M$, dvs punktskattningen plus/minus en marginal M.

(Undantag: KI för σ^2 då observationerna $X_1, \ldots, X_n$ alla är $N(\mu, \sigma)$ med okänt μ ges av $\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$, där χ^2_α är chi-två-fördelningens α-kvantil.)

För ett stickprov har vi (z och t är normalfördelningen resp. t-fördelningens kvantiler):

θ	$\hat{\theta}$	M	Villkor
μ	$\bar{X}$	$z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$	(1): σ känd; $n \geq 30$ eller N-obs.*
μ	$\bar{X}$	$t_{\alpha/2}^{(n-1)}\frac{S}{\sqrt{n}}$	(2): σ okänd & N-obs.*
p	$\hat{p}$	$z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	(3): $n \cdot \min\{\hat{p}, 1-\hat{p}\} \geq 5$.

Om man jämför två oberoende stickprov så får man liknande KI som ovan.

θ	$\hat{\theta}$	M	Villkor†
$\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2$	$z_{\alpha/2}\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	(1)‡
$\mu_1 - \mu_2$	$\bar{X}_1 - \bar{X}_2$	$t_{\alpha/2}^{(n_1+n_2-2)} S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$	(2)‡§
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1}+\frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	(3)

Det finns en signifikant skillnad mellan grupperna om nollan inte ligger i KI:t.

5 Kombinatorik

En talföljd $a_0, a_1, \ldots$ har **genererande funktion** (gf) $A(x) = \sum_{k=0}^\infty a_k x^k$, vilken kan bl.a. användas för att beräkna antalet heltalslösningar $(y_1, \ldots, y_N)$ till $\sum_{k=1}^N c_k y_k = n$ med bivillkor, och att finna slutna uttryck för talföljder som definieras med en rekursion $a_n = \sum_{k=1}^N c_k a_{n-k}$.

a_k	a^k	$\binom{n}{k}$	$\binom{n+k-1}{k}$
$A(x)$	$\frac{1}{1-ax}$	$(1+x)^n$	$\frac{1}{(1-x)^n}$

Givet $A(x)$ kan man få fram talföljden så här: $a_k = A^{(k)}(0)/k!$.

6 Varning

Detta är bara en översikt, annat ingår också i kursen, t.ex. att kunna härleda vissa resultat och också att kunna *tillämpa* sina kunskaper på praktiska problem.

Svara **aldrig** med negativa sannolikheter, sannolikheter större än ett eller negativa varianser!

*Alla observationer $X_1, \ldots, X_n$ är var för sig normalfördelade.
†(1), (2) resp. (3) skall gälla för båda stickproven.
‡Observationerna i båda stickproven måste ha samma varians.
§$S_p^2 = \frac{(n_1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$.