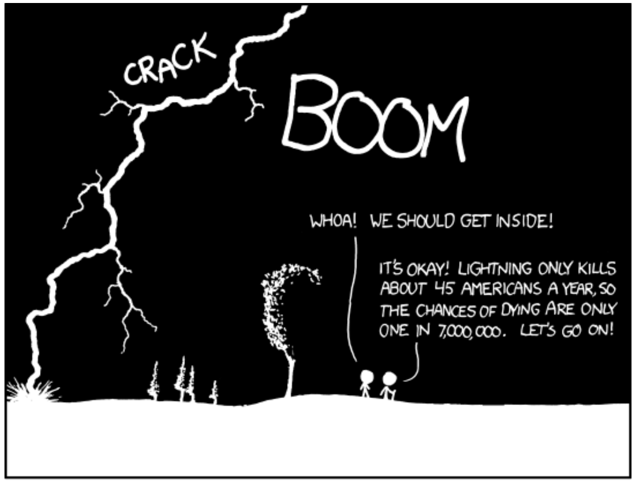
Föreläsning 3: Diskreta fördelningar

Matematisk statistik

David Bolin
Chalmers University of Technology
September 10, 2018

XKCD förklarar betingad sannolikhet



THE ANNUAL DEATH RATE AMONG PEOPLE WHO KNOW THAT STATISTIC IS ONE IN SIX.

Betingade sannolikheter

Vi vill ibland beräkna sannolikheten för en händelse A givet att vi vet att en annan händelse B har inträffat. Vi vill då veta den så kallade betingade sannolikheten för A givet B.

Betingad sannolikhet

Antag att $\mathsf{P}(B) > 0$. Den betingade sannolikheten för A givet B definieras som

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)}. \tag{1}$$

Multiplikationssatsen för sannolikheter

Om A och B är händelser så gäller att

$$\mathsf{P}(A \cap B) = \mathsf{P}(B|A)\mathsf{P}(A) = \mathsf{P}(A|B)\mathsf{P}(B).$$

Multiplikationssatsen är speciellt användbar för att beräkna sannolikheten för successiva händelser som påverkar varandra.

Bayes formel

Bayes formel

För två händelser A och B gäller att

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(B|A)\mathsf{P}(A)}{\mathsf{P}(B)}$$

Man har ofta nytta av att skriva om nämnaren $\mathsf{P}(B)$ som

$$\mathsf{P}(B) = \mathsf{P}(B|A)\mathsf{P}(A) + \mathsf{P}(B|A^c)\mathsf{P}(A^c)$$

Bayes sats lite mer generell

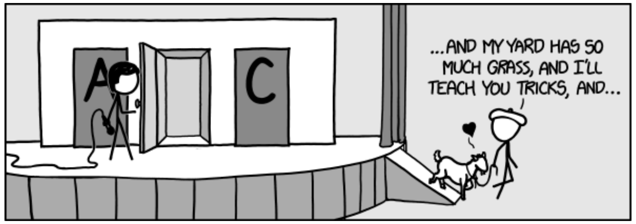
Antag $A_1, \ldots, A_n$ är disjunkta händelser så att $\Omega = A_1 \cup \cdots \cup A_n$. Låt B vara en händelse med $\mathsf{P}(B) > 0$, vi har då för $1 \leq j \leq n$

$$\mathsf{P}(A_j|B) = \frac{\mathsf{P}(B|A_j)\mathsf{P}(A_j)}{\sum_{i=1}^n \mathsf{P}(B|A_i)\mathsf{P}(A_i)}$$

The Monty Hall Problem

Ett problem baserat på TV-programmet Let's make a deal.

- I studion finns tre dörrar, bakom en dörr finns en bil och bakom de andra två finns getter.
- Den tävlande väljer en dörr utan att öppna den och programledaren öppnar en av de andra som innehåller en get.
- Den tävlande ges nu möjligheten att antingen öppna den valda dörren eller att byta och öppna den andra stängda dörren.
- Ska den tävlande byta dörr?
- Om den tävlande vill ha geten är valet enkelt, som XKCD visar:



Slumpvariabler

Ofta vill vi beskriva utfallet av ett slumpmässigt försök som ett numeriskt värde och man kan då beskriva försöket med en slumpvariabel

Slumpvariabel

En stokastisk variabel (slumpvariabel) X är en funktion som tar element från Ω som argument och som returnerar ett reellt tal.

Vi kommer beteckna slumpvariabler med stora bokstäver, ofta X, Y och Z.

Diskreta slumpvariabler

En diskret slumpvariabel kan endast anta ett ändligt eller uppräkneligt antal värden, ofta någon delmängd av heltalen.

Oberoende händelser

Två händelser sägs vara statistiskt oberoende om sannolikheten för den ena inte påverkas av om den andra inträffat eller inte. Det vill säga att $\mathsf{P}(A|B) = \mathsf{P}(A)$.

Oberoende händelser

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

Oberoende händelser

Låt $A_1, \ldots, A_n$ vara n händelser. Dessa sägs vara oberoende om vi för varje delmängd $A_{(1)}, \ldots, A_{(m)}$ har att

$$\mathsf{P}(A_{(1)} \cap A_{(2)} \cap \ldots \cap A_{(m)}) = \mathsf{P}(A_{(1)})\mathsf{P}(A_{(2)}) \cdots \mathsf{P}(A_{(m)}).$$

Sannolikhetsfunktionen

Sannolikhetsfunktion

Till en diskret stokastisk variabel X definierar vi sannolikhetsfunktionen $f(k)$ genom $f(k) = \mathsf{P}(X = k)$.

Alla funktioner är inte sannolikhetsfunktioner. Eftersom de beskriver sannolikheter måste vi ha

- $f(k) \geq 0$ för alla k.
- $\sum_{k=-\infty}^{\infty} f(k) = 1$.

Dessa två villkor är både nödvändiga och tillräckliga för att $f(k)$ ska vara en sannolikhetsfunktion. Vi har också att

$$\mathsf{P}(m \leq X \leq n) = \sum_{k=m}^{n} f(k)$$

om m och n är heltal.

Fördelningsfunktionen

Fördelningsfunktionen

Låt X vara en diskret slumpvariabel. Dess fördelningsfunktion ges då av

$$F(x) = \mathsf{P}(X \leq x) = \sum_{k \leq x} f_X(k),$$

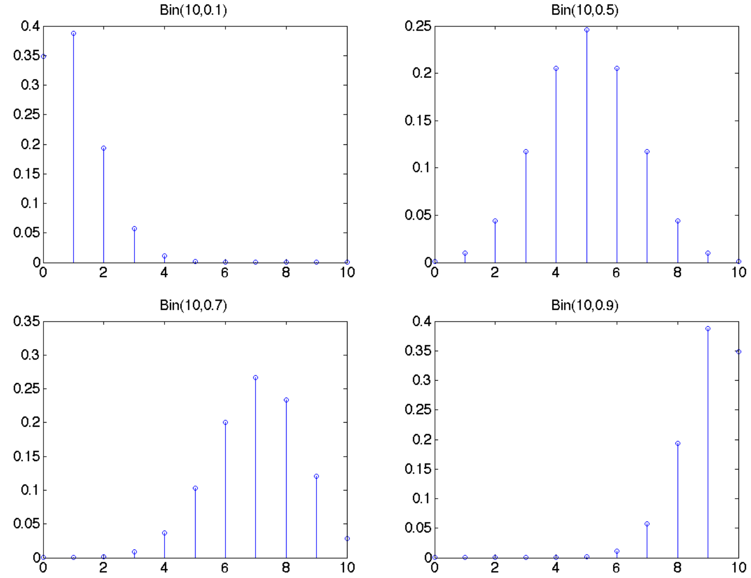
För $F(x)$ gäller att

- $F(x)$ är växande.
- $F(x) \to 1$ då $x \to \infty$.
- $F(x) \to 0$ då $x \to -\infty$.

Vi har också att:

- $\mathsf{P}(a < X \leq b) = F(b) - F(a)$.
- $\mathsf{P}(X > a) = 1 - F(a)$.
- För diskreta slumpvariabler gäller $f(m) = F(m) - F(m-1)$.

Binomialfördelningen



Bernoullifördelningen

Antag att vi gör ett försök som har sannolikhet p att lyckas och sätter

$$X = \begin{cases} 1, & \text{om försöket lyckas} \\ 0, & \text{om försöket misslyckas.} \end{cases}$$

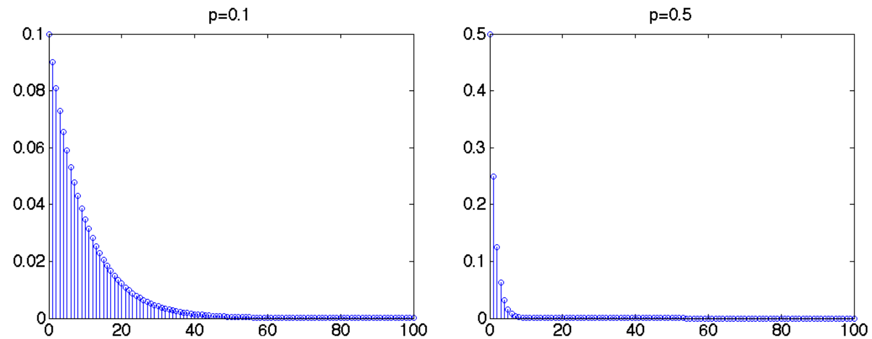
Vi har då $f(1) = p$ och $f(0) = 1 - p$.

Lite mer kompakt kan vi skriva $f(k) = p^k(1-p)^{1-k}$.

Bernoullifördelningen

En stokastisk variabel X sägs vara Bernoullifördelad om den har sannolikhetsfunktion $f(k) = p^k(1-p)^{1-k}$, k=0,1. Vi betecknar den som Be(p).

Den geometriska fördelningen



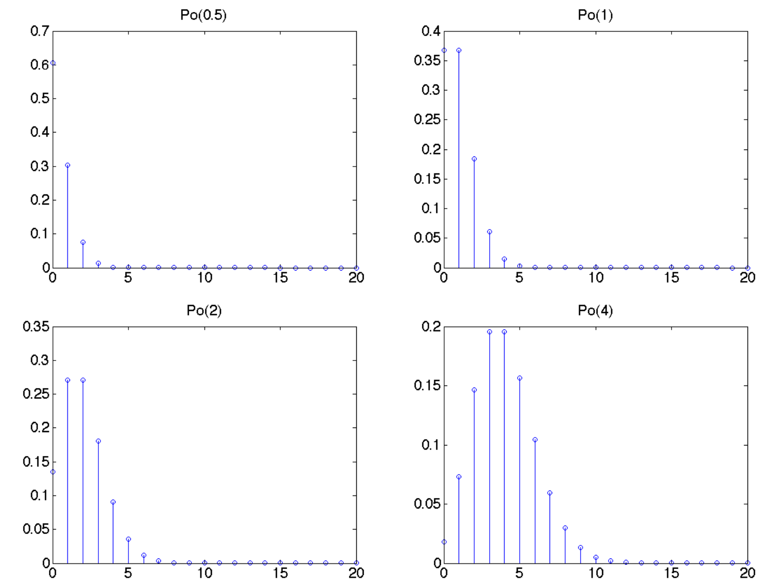
Poissonfördelningen

Poissonfördelningen är döpt efter Simeon Denis Poisson (1781-1840). Den används ofta för att beskriva fördelningen av antalet händelser av något slag som inträffar under ett tidsintervall, på en yta, eller i en volym.

Några exempel där denna fördelning passar bra är

- Räkna antalet radioaktiva partiklar som emitteras under en minut från ett radioaktivt material.
- Registrera antalet inkommande telefonsamtal till en telefonväxel under en timme.

Poissonfördelningen



Poissonfördelningen

Poissonfördelningen

En slumpvariabel X sägs vara Poissonfördelad med parameter μ om den har sannolikhetsfunktion

$$f(k) = \frac{\mu^k e^{-\mu}}{k!}$$

Vi använder ofta beteckningen Po(μ) för denna fördelning.

Summering av Poissonvariabler

Vi har vi att om X_1 är Po(μ_1) och X_2 är Po(μ_2) så gäller att $X_1 + X_2$ är Po($\mu_1 + \mu_2$).

Poissonfördelningen som gränsfördelning

Konceptuellt kan fördelningen ses som en gränsfördelning till binomialfördelningen om vi låter n gå mot oändligheten och p gå mot noll. Mer specifikt har vi

Sats

Då $n \to \infty$, $p \to 0$, och $np \to \mu$ så gäller för ett fixt heltal $k \geq 0$ att

$$\binom{n}{k} p^k (1-p)^{n-k} \to \frac{\mu^k e^{-\mu}}{k!} \tag{2}$$

I exemplen ovan kan vi tolka den här satsen som

- Det finns ett stort antal atomer (n) i materialet och sannolikheten (p) att en specifik atom ska emittera en partikel under just den tidsperioden är mycket liten.
- Ett stort antal personer kan, oberoende av varandra, ringa växeln, men för en specifik person är sannolikheten att denne ska ringa växeln mycket liten.