

Sport är en verksamhet som genererar mängder av numerisk information som följs med stort intresse. EM i fotboll är inget undantag och detta dokument visar några grafer med kommentarer om processen 'antal mål'. Resultatet beskrivs med hjälp av en Poissonfördelning eftersom 'mål' kan beskrivas som 'händelser per tidsenhet'.

(På hemsidan www.indstat.se under knappen [Arkiv] finns 'StaM-Bladet', en enkel skrift som gavs ut av SFK-StaM (Svenska Förbundet för Kvalitet, Statistisk Metodik). Nummer 1 och nummer 9 diskuterar på ett liknande sätt data från VM 1990 och VM 1994.)

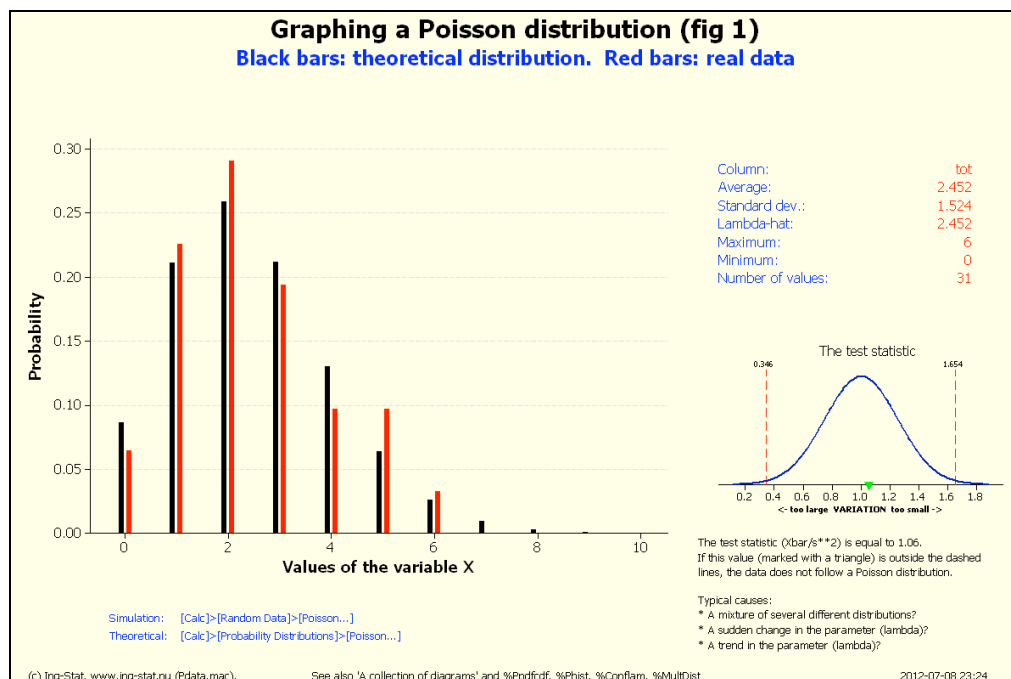
Nedan redovisas resultatet med hjälp av ett antal olika diagram (pkt 1-6):

1. Histogram över antal mål i en match
2. Histogram över tid mellan mål
3. En 'Pure birth'-process
4. Alla mål på ett 90-minutersintervall
5. Oavgjort
6. Simulering – diagram över medelvärde mot varians
7. Sammanfattning

1. Histogram över antal mål i en match

Nedanstående histogram visar totalt antal per match, dvs summan av de två lagens mål under 2x45 minuter. Förlängning och straffar är inte medräknade. Eftersom summan består av två termer, båda antages vara Poissonfördelade, blir summan också Poissonfördelad. (Om man adderar fler och fler termer kommer summan att närma sig en normalfördelning eller egentligen en Poissonfördelning som blir mer symmetrisk och klockformad).

Som en bästa skattning av fördelningens parameter (lambda) användes totalmedelvärdet och histogrammet visar verkligt utfall och den teoretiska sannolikhetsfördelningen baserad på det beräknade lambda:

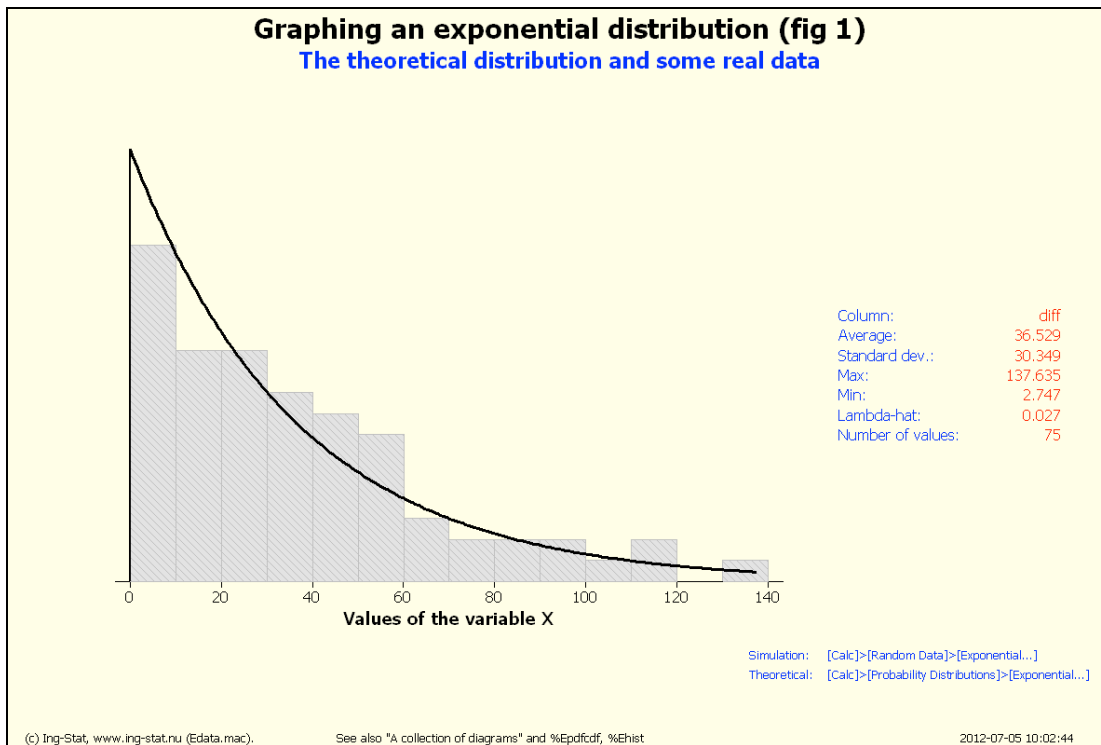


Det verkar vara en god anpassning mellan verkligt och teoretiskt utfall.

För en Poissonfördelning gäller att förhållandet mellan väntevärde och varians är 1. För fotbollsresultatet är denna siffra 1.06 (grön markering), alltså en god överensstämmelse.

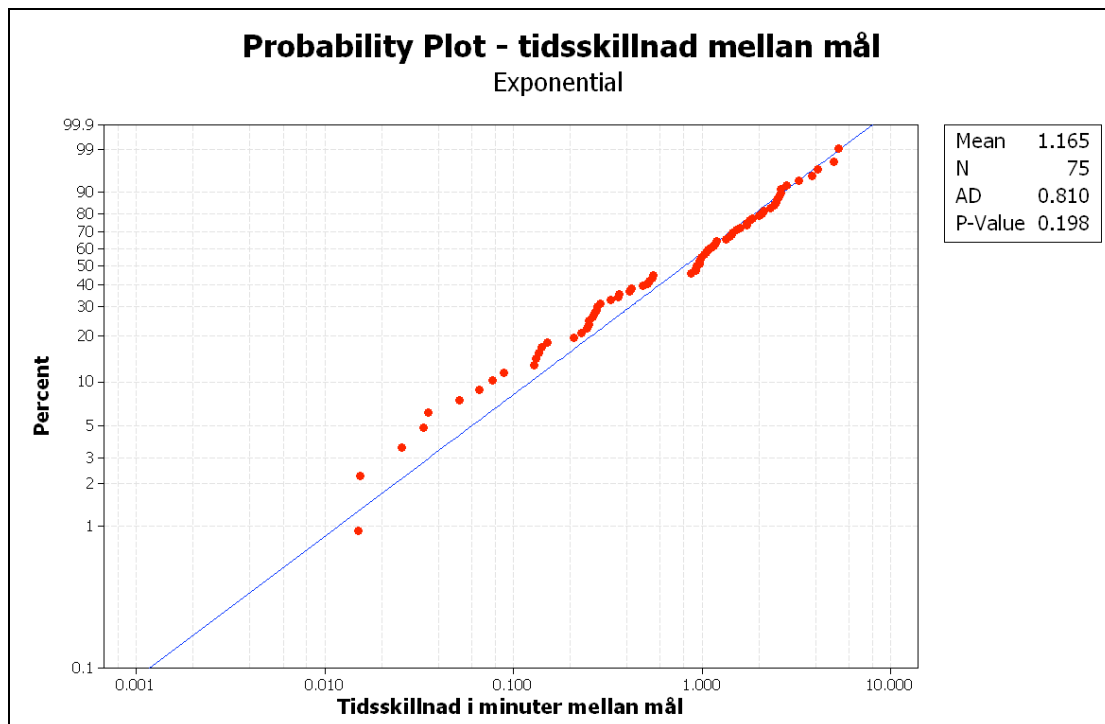
2. Histogram över tid mellan mål

Ett annat sätt att titta på en händelseprocess i tiden är att beräkna tidsavståndet mellan händelserna. Om händelser uppträder enligt en Poissonprocess på tidsaxeln så är tidsavstånden mellan händelserna exponentialfördelade. Nedanstående histogram visar dessa tidsavstånd och eftersom det fanns 76 mål i EM 2012 så finns det 75 tidsavstånd. Teoretisk gäller för en exponentialfördelning att väntevärde och standardavvikelse är exakt samma. I detta utfall blev resultatet 36.5 respektive 30.3:



Det verkar vara en god anpassning mellan verkligt och teoretiskt utfall.

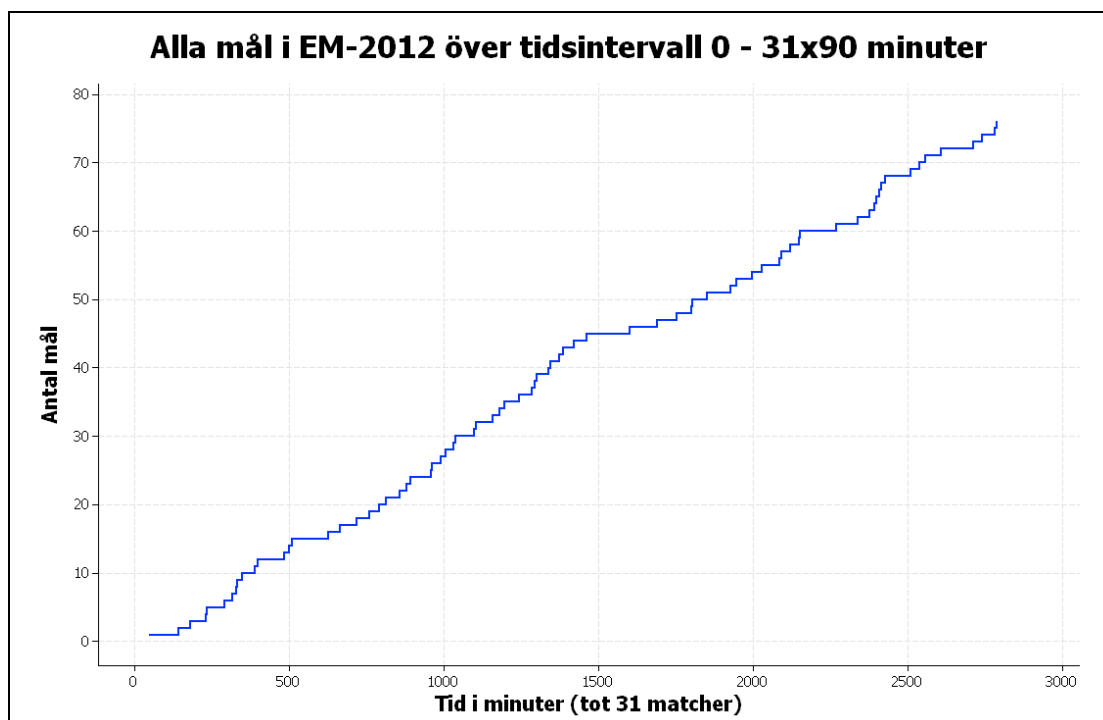
Om man ritat resultatet som en vanlig 'probability plot' får man följande resultat (minuter):



3. En 'Pure birth'-process

Inom teorin för stokastiska processer skulle 'mål i en fotbollsmatch' betraktas som en 'pure birth'-process där varje mål alltså är en 'birth'. (In en 'Birth and death'-process sker det 'händelser in' och 'händelser ut' t.ex. nya felrapporter in och åtgärdade felrapporter ut.)

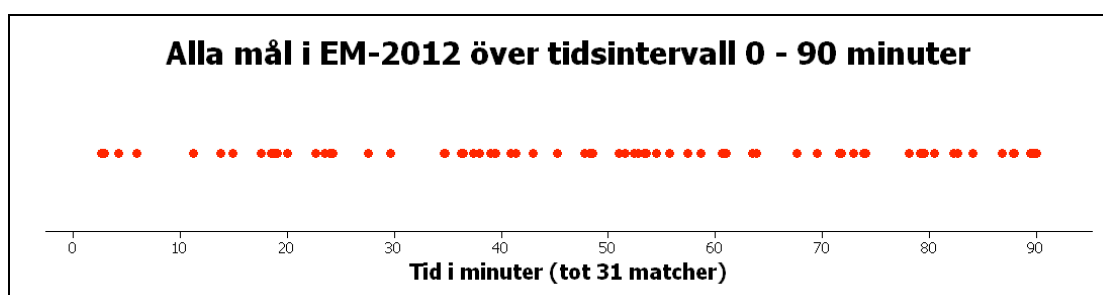
Om man ritat alla matchernas resultat i en lång tidserie (31 x 90 min) får man följande diagram där lutningen är den genomsnittliga målintensiteten. För varje mål blir det ett steg uppåt på Y-axeln:



Diagrammet visar antal mål som en 'pure birth'-process.

4. Alla under 90 min

Om man ritat varje mål som en händelse på tidsaxeln 0 – 90 minuter får man följande diagram. Här skulle man kanske kunna se uttunning av processen då det närmar sig halvtidsvila eller slutet på matchen men ingen tydlig sådan effekt kan skönjas:



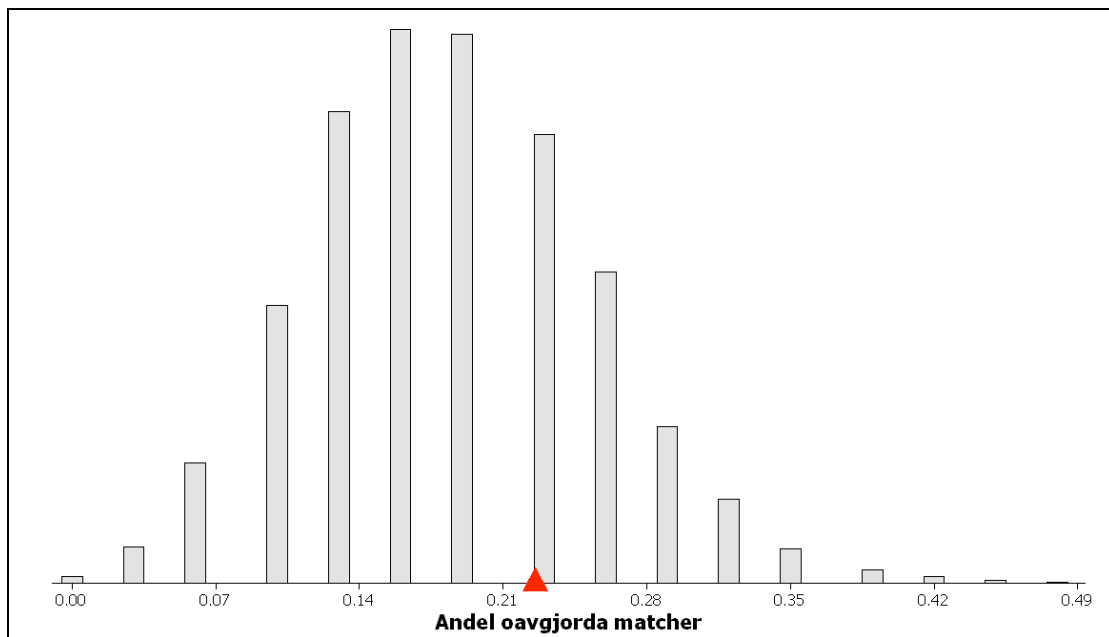
5. Oavgjort

Givet ett visst värde på parametern λ och att λ är konstant mellan lagen, kan vi beräkna sannolikheten för en oavgjord match. Detta är i grunden en enkel sannolikhetsberäkning. (I tabellen nedan sker beräkningen t.o.m. resultatet 10 - 10, högre är väldigt osannolikt.):

Resultat	Sannolikhet
0 - 0	$0.0863 \cdot 0.0863 = 0.0074$
1 - 1	$0.2114 \cdot 0.2114 = 0.0447$
2 - 2	$0.2590 \cdot 0.2590 = 0.0671$
.	
.	
.	summa: 0.185

Sannolikheten att en fotbollsmatch slutar i oavgjort är alltså 0.185 (baserat på att 'antal mål' följer en Poissonfördelning med parametern 2.45). Resultatet från de 31 matcherna i EM 2012 visar att det blev 7 matcher som blev oavgjorda. Detta ger proportionen $7/31 = 0.223$. Nu är det en viss skillnad på 0.185 och det verkliga utfallet 0.223 men är det en systematisk eller en slumpmässig skillnad?

Vi kan göra en enkel hypotesprövning men kanske en simulering kan ge en bättre beskrivning. Vi har alltså en situation vars utfall kan bäst beskrivas med binomialfördelningen där $n = 31$ (fotbollsmatcher) och $p = 0.185$. Nedan finns ett histogram över resultatet från 10000 partier. Varje resultat har dividerats med 31 för att få den observerade felkvoten:

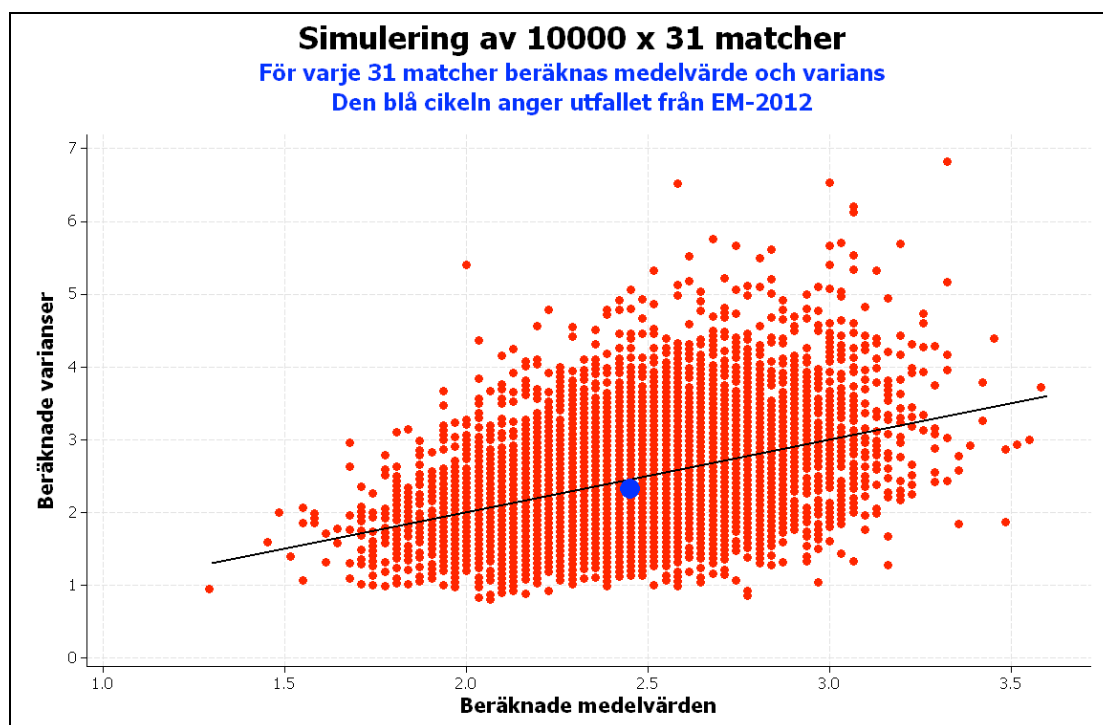


Simulering av 10000 partier om $n = 31$ och $p = 0.185$. (10000 partier kan alltså ses som 10000 EM-slutspel.) Den röda triangeln visar resultatet från EM-2012 och det är inte speciellt avvikande.

Det finns knappast underlag att påstå att resultatet 0.223 avviker signifikant, dvs nollhypotesen förkastas inte.

6. Simulering – diagram över medelvärde mot varians

Om man simulerar ett stort antal EM-turneringar kan man plotta medelvärdena mot varianserna:



Simulering av 10000 EM-turneringar med $\lambda = 2.45$. För varje 'turnering' beräknades medelvärde och varians. Den svarta linjen går genom punkterna $y = x$ dvs att teoretiskt är väntevärdet lika med variansen i en Poissonfördelning. Observera att fördelningen av variansen är (typiskt) en aning skev - för ett givet medelvärde är det fler punkter ovan linjen än under linjen.

7. Sammanfattning

Detta dokument visar några sätt att med olika diagram åskådliggöra resultaten från EM-fotbollen 2012. Det genomsnittliga antal mål blev 2.45 (för VM-fotbollen 1990 respektive 1994 blev motsvarande värde 2.21 respektive 2.67.)

I ovanstående framställning har λ betraktats som en konstant. I mer seriösa och kanske mer komplicerade fall kan man ansätta att λ följer en slumpmässig fördelning. Man menar då, applicerat på fotboll, att lagens λ kan anses vara dragna från en definierad sannolikhetsfördelning som man valt på olika grunder.

Man skulle också kunna diskutera spelarnas individuella λ (något som man gärna gör på sportsidorna), men summan av dessa blir ju dock lagets λ ... ■

Ingemar Sjöström