

Sportssystemet

Revisjon av sportrekningsalgoritmen i sportssystemet

Versjon 1.0

Lars Erik Gabestad

Sjefskonsulent Capgemini

1. Innledning

Sportssystemet er en omfattende applikasjon der blant annet sporttrekk foregår. I sportssystemet kan dette gjøres både maskinelt og manuelt, men maskinell, tilfeldig sporttrekk er det som i all hovedsak brukes. Styret i DNT ønsker å kvalitetssikre det maskinelle sporttrekket og Capgemini ble engasjert for å verifisere at det er vilkårlig.

2. Metode

Den benyttede funksjon for å sikre et vilkårlig sporttrekk, heretter kalt ”Den vilkårlige funksjon”, er levert av Oracle og følger med som standard i databasen.

Verifikasjon av det vilkårlige sporttrekket foregikk ved å undersøke benyttede programmeringsalgoritmer og hvordan disse er blitt benyttet i Sportssystemet.

Det ble foretatt uavhengige tester av den vilkårlige funksjon, samt tester gjennom Sportssystemet. Dette ble gjort for å finne den vilkårlige funksjons vilkårlighetsgrad.

For å avgjøre om sporttrekket er vilkårlig har vi tatt utgangspunkt i hvor mange ganger hvert enkelt mulig tilfelle har slått til (Den relative frekvensen), og sammenlignet den med den beregnede sannsynligheten.

Et eksempel:

Et løp har ti startende og samme distanse. Hver startende vil da ha 1/10 (eller 0,1=10%) sjanse for å få for eksempel spor 1. For å verifisere at sporttrekket er vilkårlig, skal den relative frekvensen nærme seg 1/10 etter hvert som antall trekninger øker.

Samme type tester kjøres mot den vilkårlige funksjon uavhengig av Sportssystemet. Dette gjøres for å teste funksjonen alene og verifisere at den er vilkårlig.

3. Konklusjon

Maskinelt sporttrekk er ut i fra våre tester og vurderinger vilkårlig.

4. Grunnlag

Grunnlaget for konklusjonen blir her presentert grundigere. Vi har utført 100000 sporttrekk på fem forskjellige løp, og 250000 sporttrekk på et løp. Løpene det ble testet mot har forskjellig antall startende og forskjellige distanser. I alle disse tilfellene har vi sett at sannsynligheten for en bestemt startende å få et bestemt spor blir tilnærmet lik ved gjentatte trekninger. Den relative hyppigheten nærmer seg den forventede hyppigheten (beregnete sannsynligheten) ved mange trekninger, og avviket blir mindre ved mange trekninger. Dette gjelder både der alle har samme distanse, eller løp der det er tillegg.

Sportssystemet bruker den vilkårlige funksjon uavhengig av data i Sportssystemet. Det betyr at verdier i databasen lagret av Sportssystemet (som for eksempel hest_id), ikke har noen innvirkning på sporvalget. Unntaket til denne regelen er distansen for hver startende. Ved sporttrekk skal de med samme distanse få etterfølgende spor. Det betyr at for eksempel spor 1 til 4 har samme distanse, 5-8 en annen distanse osv. Derfor er det korrekt at distansen skal påvirke sporvalget slik at de med samme distanse får etterfølgende spor.

Vi presenterer her trekninger fra tre løp, og en uavhengig test av den vilkårlige funksjonen.

4.1 12 startende med samme distanse.

Figur 1 illustrerer sporttrekk for et løp med 12 startende på distansen 1609 meter. Vi utførte sporttrekk 250000 ganger. Her følger en liten forklaring til figuren.

- Den grønne linjen, Sannsynlighet, beregnet sannsynligheten. I dette tilfelle med 12 startende på samme distanse blir det $1/12$, som er tilnærmet 0,0833.
- Blå linje, Max rh, er største relative frekvens (antall forekomster av en spesiell startende i et spesielt spor).
- Rød linje, Min rh, er minste relative frekvens.
- Lilla linje, Avvik, er et mål for det største avviket mellom beregnet sannsynlighet og relativ hyppighet.
- Horisontal akse er antall trekninger, og vertikal akse er måltallet for sannsynlighet/frekvens.

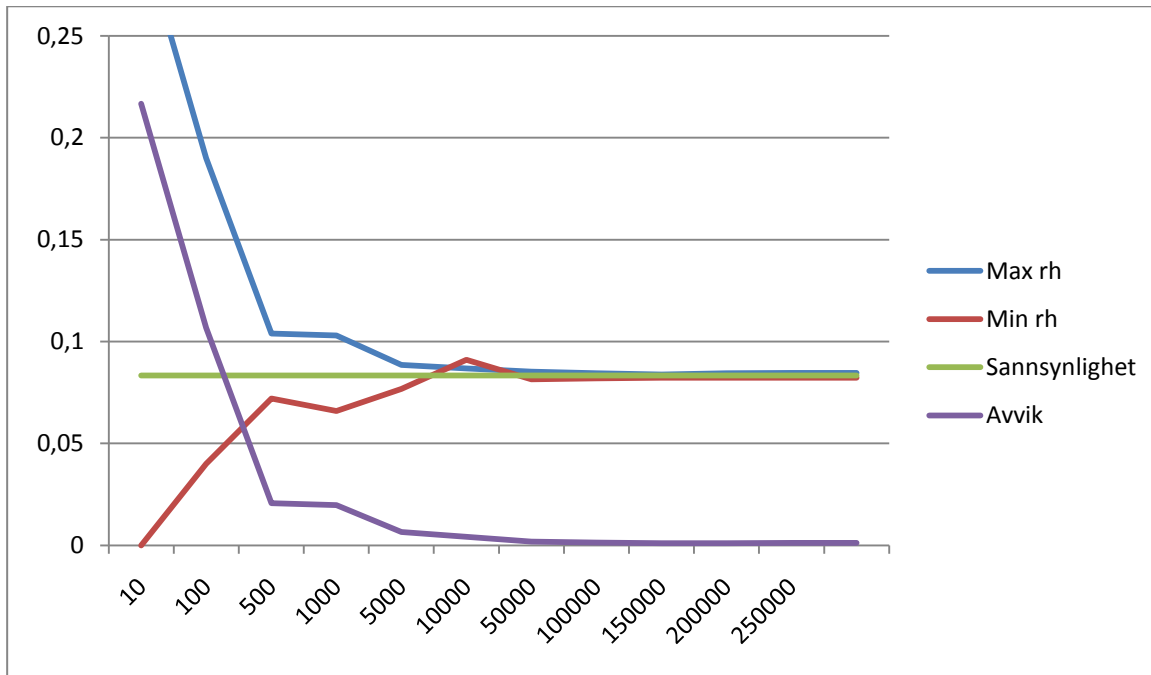


Figure 1 – 12 startende, samme distanse

Vi ser av Figur1 at Max rh og Min rh nærmer seg beregnet sannsynlighet ved mange trekninger og at avviket nærmer seg null. Bakgrunnstallene viser at ved 100 trekninger har alle startende vært innom alle spor minst en gang. Max rh er 0,19. Det betyr at en hest har samme spor maks 19 ganger. Min rh er 0,04, som betyr at det minste antallet en startende har samme spor er 4. Dette er en ganske stor variasjon, men forklares ved at det er utført få trekninger i forhold til antall startende. Nå vi kommer opp til 500 trekninger er Max rh blitt 0,104 og Min rh er 0,072. Dette nærmer seg den beregnede sannsynligheten på 0,833, og etter 250000 sportrekk er verdiene for Max rh og Min rh blitt 0,0846 og 0,0823. Avviket er da på 0,0013.

Dette betyr at etter 250000 sportrekk viser det seg at hyppigheten for å få et bestemt spor ligger på mellom 8,23% og 8,46%. Det er et maksimalt avvik fra beregnet sannsynlighet (8,33%) på 0,13 prosentpoeng.

4.2 15 startende, fire distanser

Figur 2 illustrerer sportrekk for et løp med 15 startende og fire forskjellige distanser. Figur 1 viser sporene 1-4 for de fire startende med kortest distanse. Figur 3 viser startende med 20 meter tillegg, Figur fire viser de seks startende med 40 meter tillegg og figur 4 viser de 2 startende med 60 meter tillegg. Vi utførte sportrekk 100000 ganger. Figuren inneholder samme type data som Figur 1.

2009-08-17

Revisjon av sportrekningsalgoritmen i sportssystemet

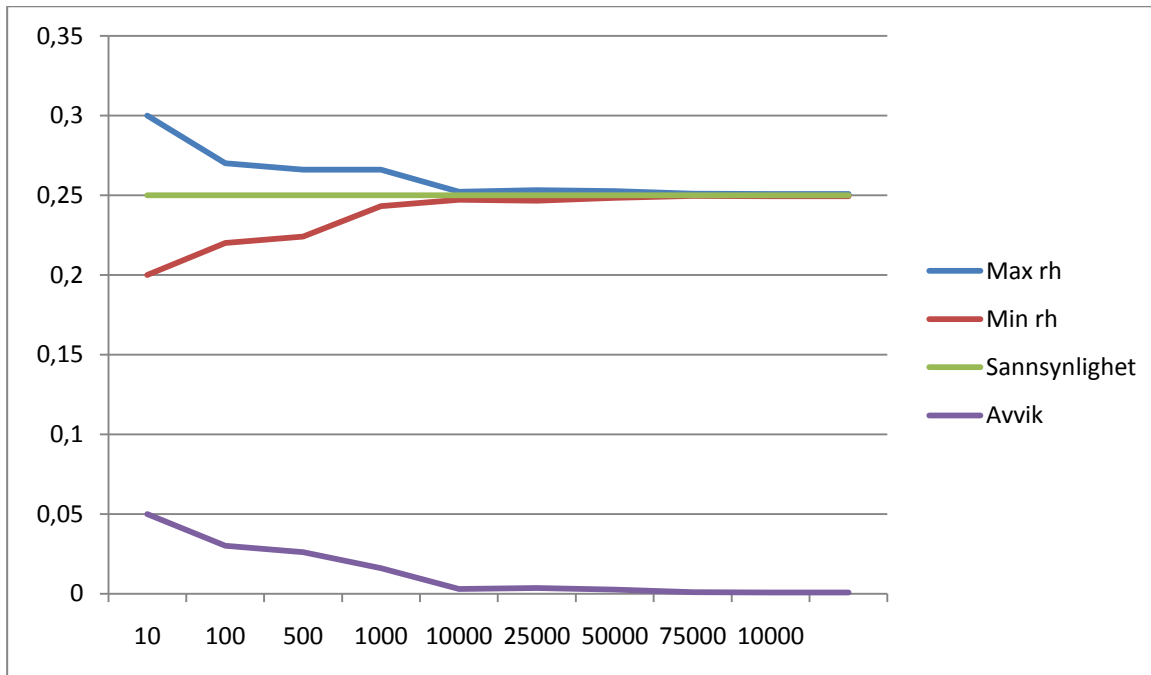


Figure 2 – 4 startende uten tillegg

Vi ser av Figur2 (fire hester med samme distanse) at Max rh og Min rh nærmer seg beregnet sannsynlighet ved mange trekninger og at avviket nærmer seg null. Bakgrunnstallene viser at allerede ved 10 trekninger har alle startende vært innom alle spor minst en gang. Max rh er her 0,3 som betyr at en hest har samme spor maks 3 ganger på disse ti trekningene. Min rh er 0,2, som betyr at det minste antallet en startende har samme spor er 2. Når vi kommer opp til 500 trekninger er Max rh blitt 0,266 og Min rh er 0,224. Dette nærmer seg den beregnede sannsynligheten på 0,25, og etter 100000 sportrekk er verdiene for Max rh og Min rh blitt 0,2508 og 0,2493. Avviket er da på 0,0008.

Dette betyr at etter 100000 sportrekk viser det seg at hyppigheten for å få et bestemt spor ligger på mellom 24,93% og 25,08%. Det er et maksimalt avvik fra beregnet sannsynlighet (25%) på 0,08 prosentpoeng.

Figur 3 viser det samme for tre hester med 20 meter tillegg. Vi ser igjen at avviket går mot null, og at den relative frekvensen går mot forventet frekvens. Etter 100000 sportrekk viser det seg at hyppigheten for å få et bestemt spor ligger på mellom 33,22% og 33,54%. Det er et maksimalt avvik fra beregnet sannsynlighet(33,33%) på 0,3 prosentpoeng.

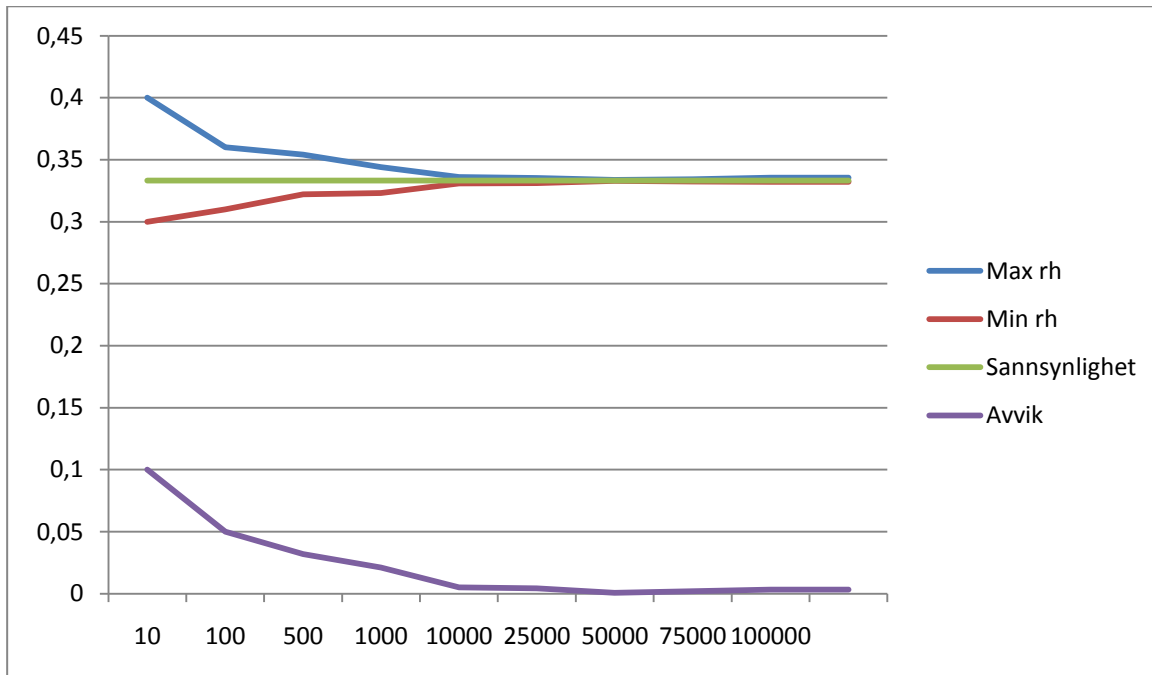


Figure 3 – Tre startende med 20 meter tillegg

Figur 4 viser hvordan data for seks hester med 40 meter tillegg fordelte seg. Her er det en litt utypisk observasjon ved at den minste relative hyppighet for 500 trekninger er nærmere forventet hyppighet enn ved 1000 trekninger. Ved 1500 trekninger er minste relative frekvensen igjen nærmere forventet frekvens. Ser vi nærmere på tallene virker det å være den "gode" verdien ved 500 trekninger som er utypisk, og ikke verdien ved 1000 trekninger. Ser vi videre på figur 4 ser vi at avviket også her går mot null, og den relative frekvensen går mot forventet frekvens. Etter 100000 sporttrekk viser det seg at hyppigheten for å få et bestemt spor ligger på mellom 16,55% og 16,85%. Det er et maksimalt avvik fra beregnet sannsynlighet(16,67%) på 0,18 prosentpoeng.

2009-08-17

Revisjon av sportrekningsalgoritmen i sportssystemet

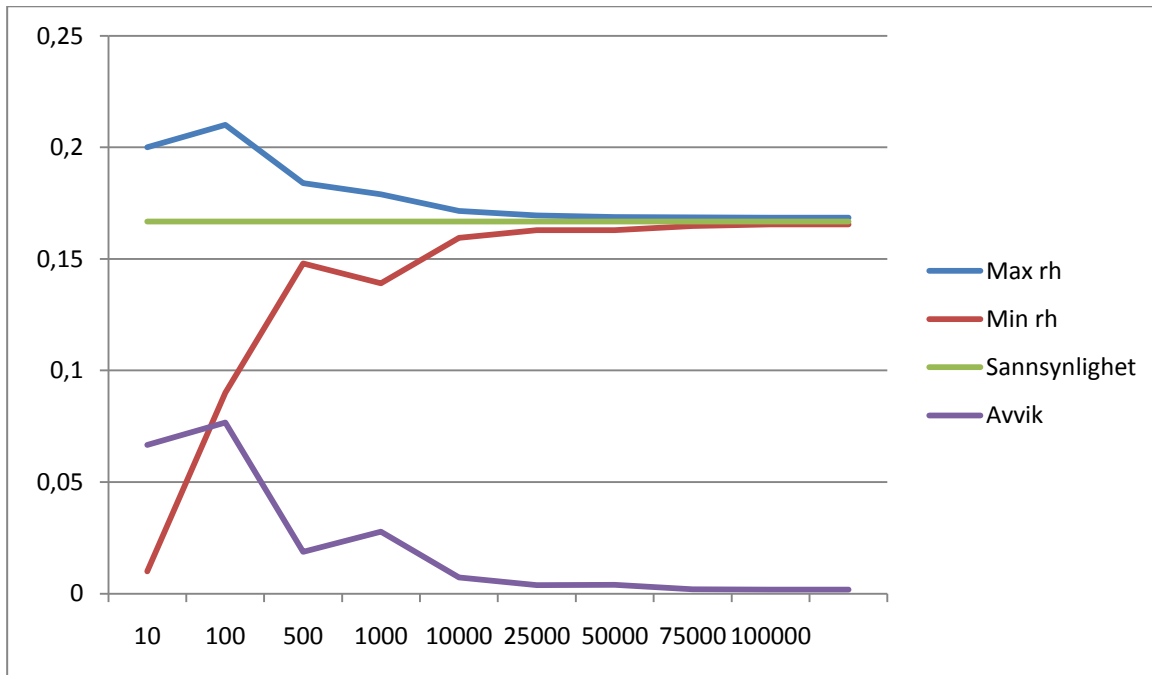


Figure 4 – Seks startende med 40 meter tillegg

Figur 5 viser hvordan data for to hester med 60 meter tillegg fordelte seg. Igjen går avviket mot null, og at den relative frekvensen går mot forventet frekvens. Etter 100000 sportrekk viser det seg at hyppigheten for å få et bestemt spor ligger på mellom 49,83% og 50,17%. Det er et maksimalt avvik fra beregnet sannsynlighet (50%) på 0,17 prosentpoeng.

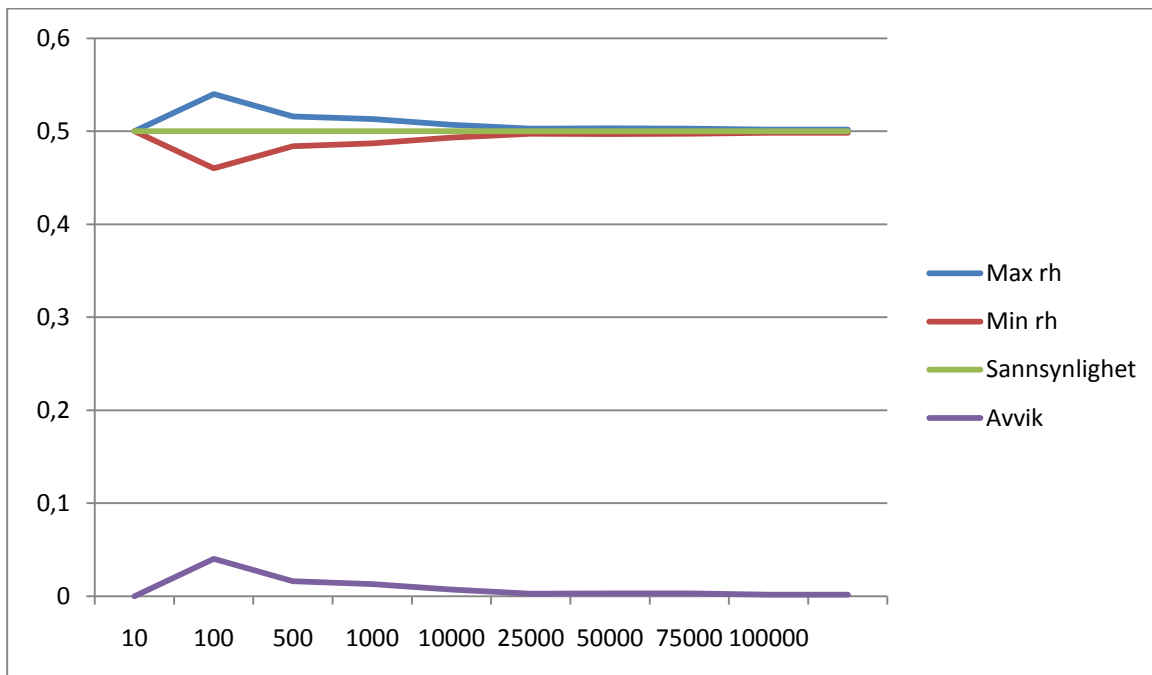


Figure 5 – To startende med 60 meter tillegg

4.3 10 startende med samme distanse.

Figur 6 illustrerer sporttrekk for et løp med 10 startende på samme distanse. Vi utførte sporttrekk 100000 ganger.

Vi ser av Figur6 at Max rh og Min rh også her nærmer seg beregnet sannsynlighet ved mange trekninger og at avviket nærmer seg null. Bakgrunnstallene viser at ved 100 trekninger har alle startende vært innom alle spor minst en gang. Max rh er her 0,13 som betyr at en hest har samme spor maks 13 ganger på disse 100 trekningene. Min rh er 0,06, som betyr at det minste antallet en startende har samme spor er 6 ganger på de første 100 trekningene. Når vi kommer opp til 10000 trekninger er Max rh blitt 0,1062 og Min rh er 0,0963. Dette nærmer seg den beregnede sannsynligheten på 0,1, og etter 100000 sporttrekk er verdiene for Max rh og Min rh blitt 0,1017 og 0,099. %. Det er et maksimalt avvik fra beregnet sannsynlighet (10%) på 0,17 prosentpoeng.

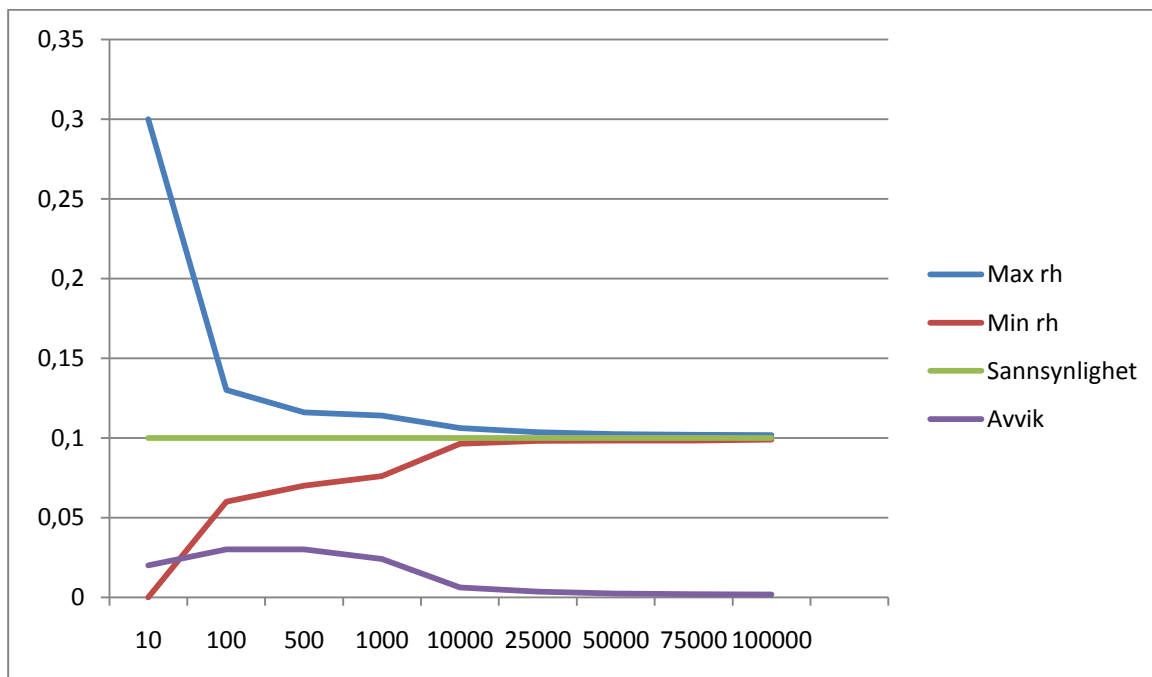
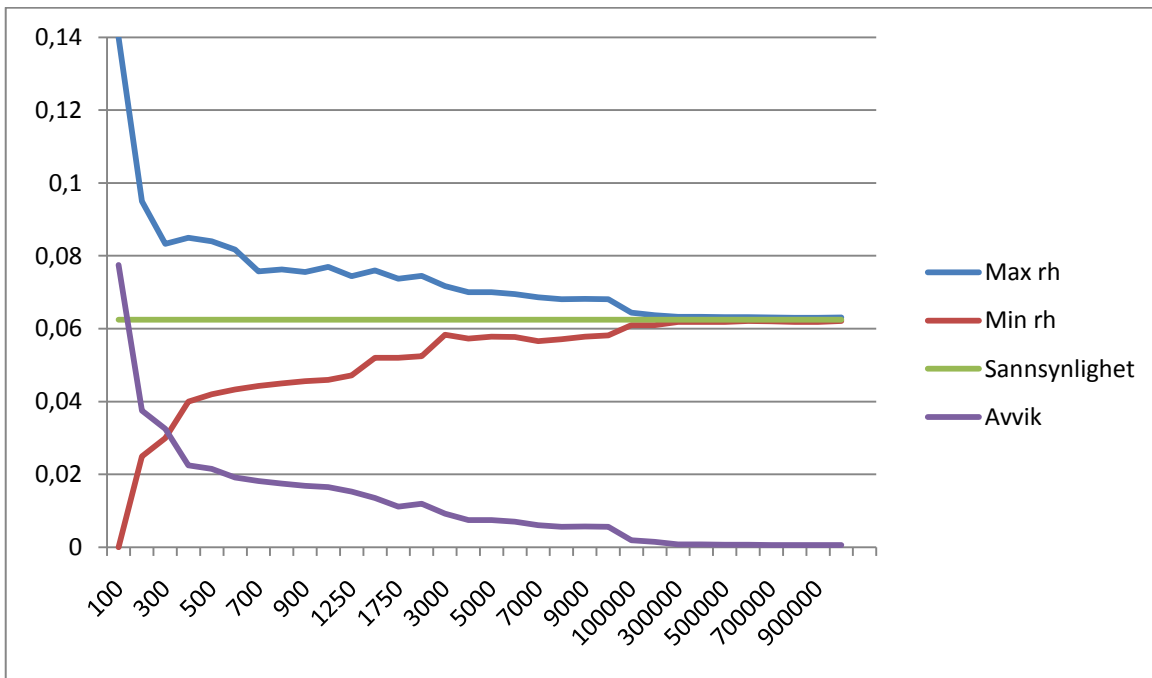


Figure 6 – 10 startende, same distance

4.4 Uavhengig test av vilkårlig funksjon

Den vilkårlige funksjonen er levert av Oracle og følger med database. Vi presenterer her en test der vi vil velge et vilkårlig tall mellom 1 og 16. Forsøket er kjørt en million ganger. Figur 7 viser hvordan relativ hyppighet utviklet seg i forhold til forventet hyppighet. Vi ser igjen at avviket nærmer seg 0 når antallet stiger. Max rh og Min rh går også sakte mot den forventede frekvensen på 1/16. Etter en million forsøk er avviket på 0,0006 prosentpoeng.



Lars Erik Gabestad (sign)

Capgemini Norge AS