Projekt 2 Rapport

Kurskod: IX1501
Datum: 2017-11-29

Evan Saboo, saboo@kth.se
Max Kufa, mkufa@kth.se

Uppgift 1: Distribution Parameters

Sammanfattning

Uppgift

Uppgift 1 består av tre deluppgifter:

- Estimera μ och σ från delvis slumpad data.
- Testa de tre följande modellerna till den slumpade datan, och välj ut den som bäst representerar den slumpade datan.

1. $\text{pdf}(x) \propto e^{-a\,x}$, $x \geq 0$
2. $\text{pdf}(x) \propto x\,e^{-a\,x}$, $x \geq 0$
3. $\text{pdf}(x) \propto x\,(x+b)\,e^{-a\,x}$, $x \geq 0$

- Beräkna ett “Confidence Interval” på 95% för de valda konstanterna i den valda modellen. Intervallet får inte avvika mer än 10% från mellanvärdet, e.g. μ (1±0.05).

Den slumpade datan genereras från en fördefinierad fil vid namn “Projekt2.m”

Resultat

Estimera μ och σ:

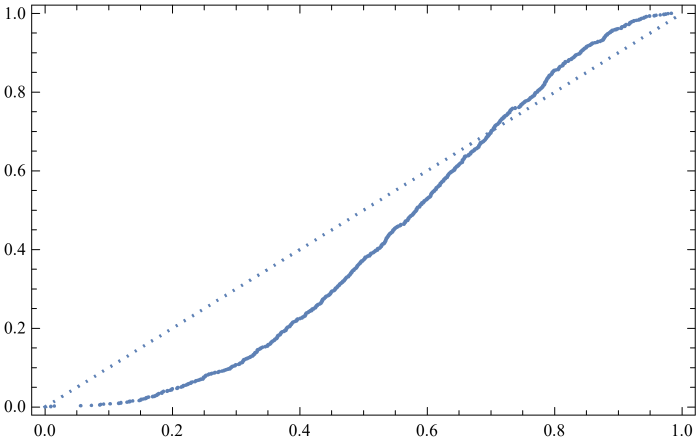
Genom att använda “Mean” och “StandardDeviation” så kan vi estimera μ och σ ur en lista av data genererad av “randomNumber” av storleken 1000:

μ = 5.19974
σ = 3.24391

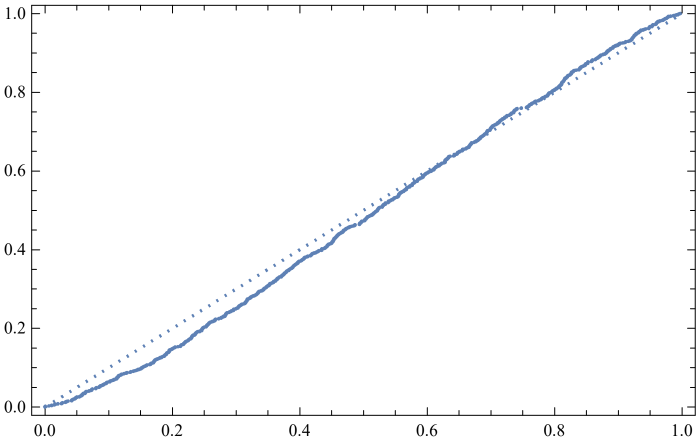
Testa de tre modellerna:

Vi använder oss av “ProbabilityDistribution” och “EstimatedDistribution” för att estimera våra matematiska funktioner som distributionsmodeller gentemot vår data. De olika kurvorna beter sig olika i jämförelse med vår datamängd $\mathbb{x}$, och det gäller att hitta den modell som best passar in på $\mathbb{x}$.
Följande grafer visar modellernas beteende i jämförelse med datamängden:

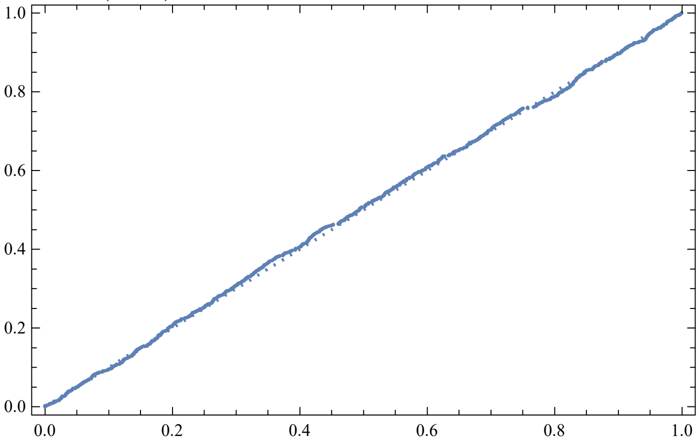
Modell e^{-a*x}



Modell $x * e^{-a*x}$



Modell $x(x+b) * e^{-a*x}$



Vi ansåg att den sista modellen, $x(x+b) * e^{-a*x}$, gav en funktion som bäst överrensstämmde med $\mathbb{x}$.
α och β fick värdena 4.32677 respektive 0.48208.

Beräkna “Confidence Interval”

Målet här är att hitta de gränsvärden för α och β som fortfarande ger ett “coinfidence interval” på 10%.
Vi bestämde oss för att använda α och β från den föregående uppgiften som hjälp för att hitta gränsvärdena, och använde metoden i sektion 2.2.3.2 i projektbeskrivningsfilen för att manuellt beräkna gränsvärderna genom att substituera μ med α och β.

Gränsvärderna för α respektive β blev därmed:
$0.464813 \leq \alpha \leq 0.53476$
$2.99036 \leq \beta \leq 3.0603$

Kod

Uppgift 2: Which Distribution?

Sammanfattning

I uppgift 2 ska vi använda fördelningarna $i = 1, 2, \ldots, 6$ i randomNumber funktionen och svara på dessa frågor:

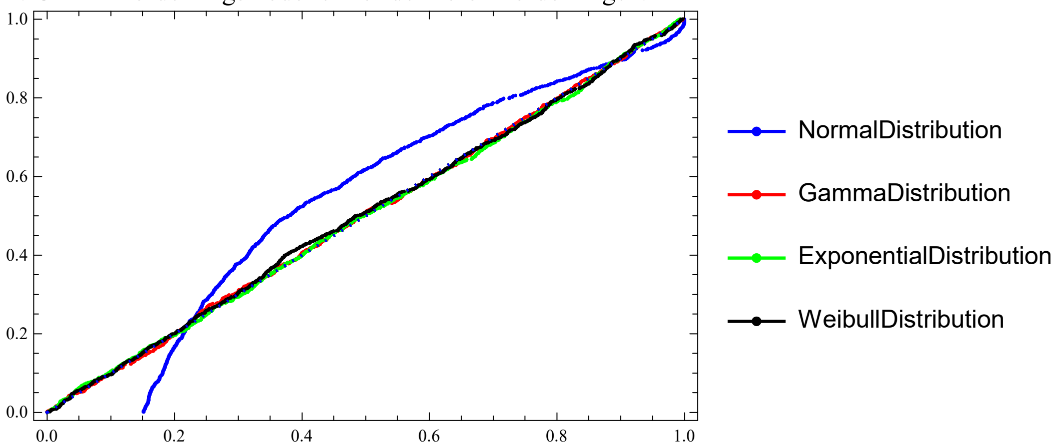
- Använd räta linjens metod för att bestäma en distributionsmodell för var och en av data-varianterna.
- Estimera parametrarna för de olika distributionsmodellerna
- Beräkna 95% konfidensintervaller för parametrarna. Intervallet bör inte vara längre än 10% av dess medelvärde.

Resultat

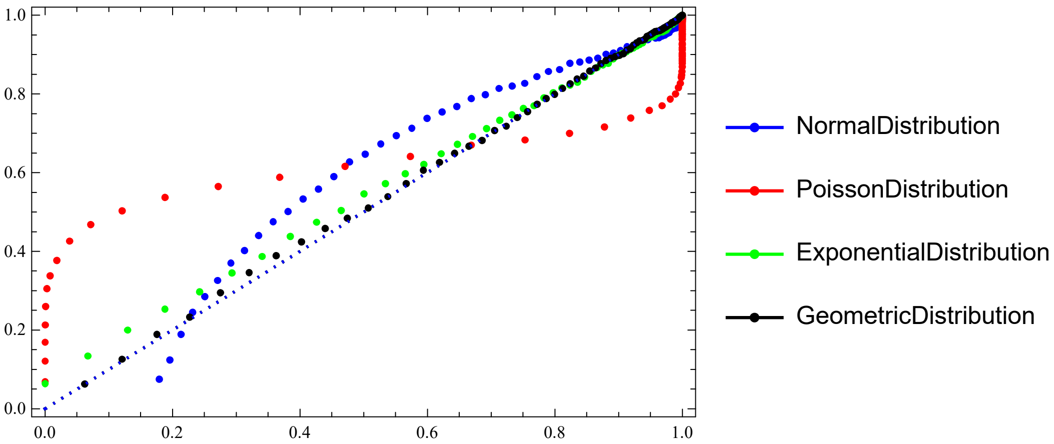
Bestäm fördelningsmodell

För att hitta den rätta fördelningsmodeller för varje fördelning i randomNumber funktionen testa vi alla inbyggda fördelningsmodeller och tog den som var närmast räta linjen.

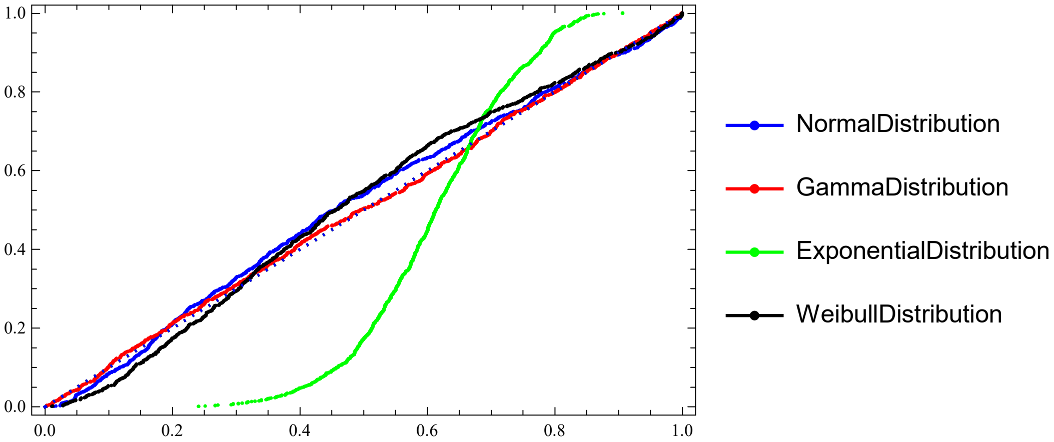
1. Gamma fördelningsmodellen för den första fördelningen.



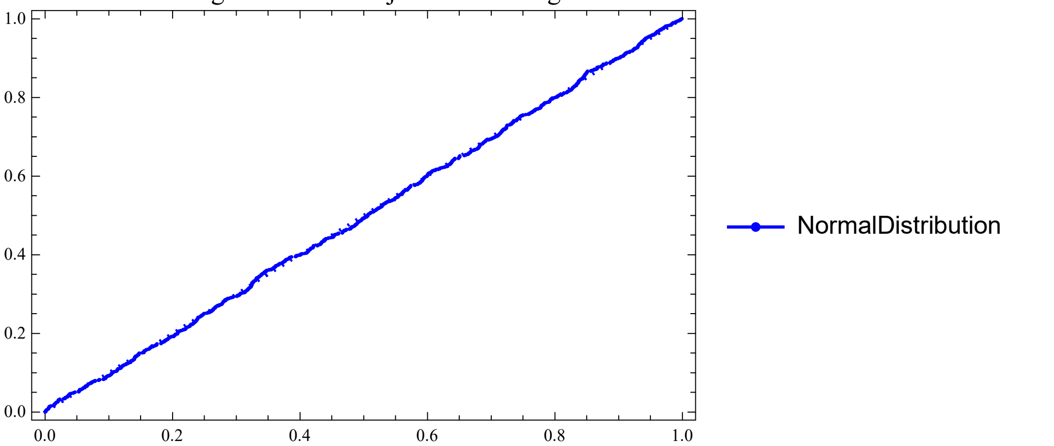
2. Geometriska fördelningsmodellen för andra fördelningen.



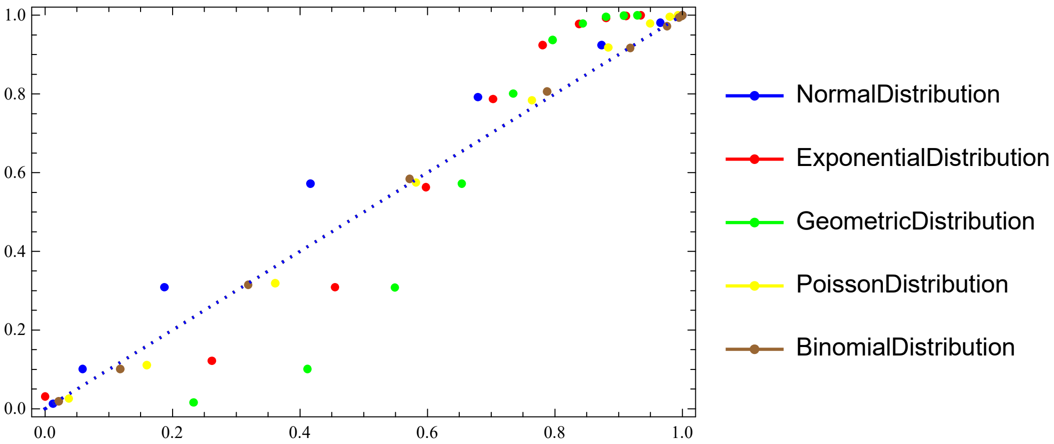
3. Gamma Distributions modellen för tredje fördelningen.



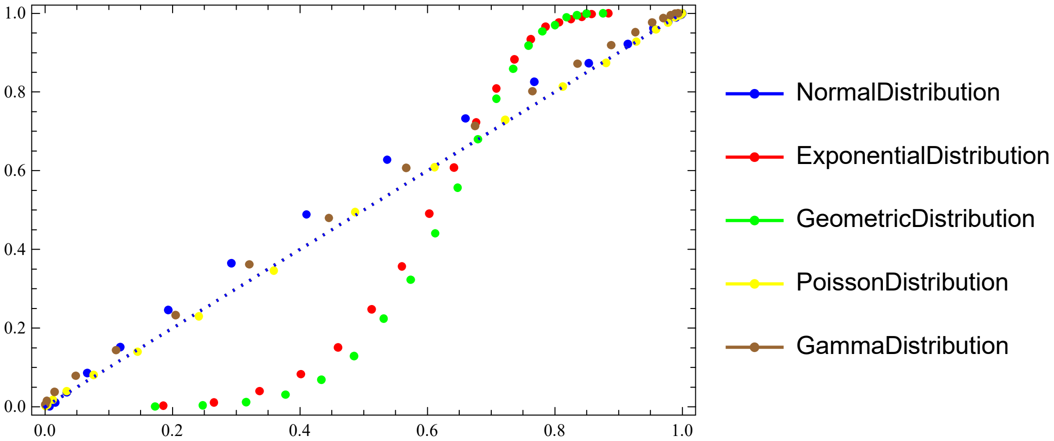
4. Normalfördelningsmodellen för fjärde fördelningen.



5. Poissonfördelningsmodellen för femte fördelningen.



6. Poissonfördelningsmodellen för sjätte fördelningen.



Estimera parametrarna

Eftersom vi fick ut fördelningsmodell för fördelning så kan vi också få estimera parametrarna med hjälpa av inbyggda funktionen “EstimateDistribution”.
Vi fick parametrarna:

Distribution 1:
`GammaDistribution[0.970294, 0.216259]`

Distribution 2:
`GeometricDistribution[0.0690512]`

Distribution 3:
`GammaDistribution[7.96105, 1.61395]`

Distribution 4:
`NormalDistribution[2.98273, 5.15498]`

Distribution 5:
`PoissonDistribution[3.275]`

Distribution 6:
`PoissonDistribution[9.758]`

Beräkna konfidensintervaller

För att få ut konfidensintervaller för varje fördelning tog först få ut $\{\alpha, \beta\}$ eller bara $\{\alpha\}$ (beroende på distributionsmodellen) med hjälp av `EstimatedDistribution` och sedan använda Central Limit Theorem för att få ut intervallerna för α och β.
Vi kom fram till:

Distribution 1:
$0.964902 < \alpha < 1.01956$
$0.213129 < \beta < 1.01956$

Distribution 2:
$0.0631177 < \alpha < 0.0654572$

Distribution 3:
$7.76802 < \alpha < 8.49892$
$1.52164 < \beta < 1.61731$

Distribution 4:
$2.96675 < \alpha < 3.15467$
$4.9226 < \beta < 5.26378$

Distribution 5:
$3.14326 < \alpha < 3.29407$

Distribution 6:
$9.54735 < \alpha < 10.0513$

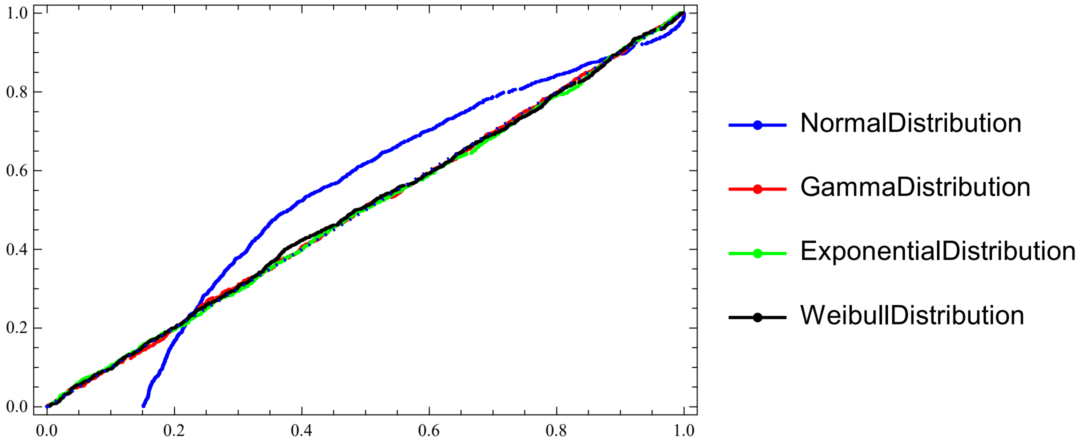
Genom att repitera uträckningen antal gånger så fick vi ut intervaller som är längre än 10% av dess medelvärde.

Uträkning

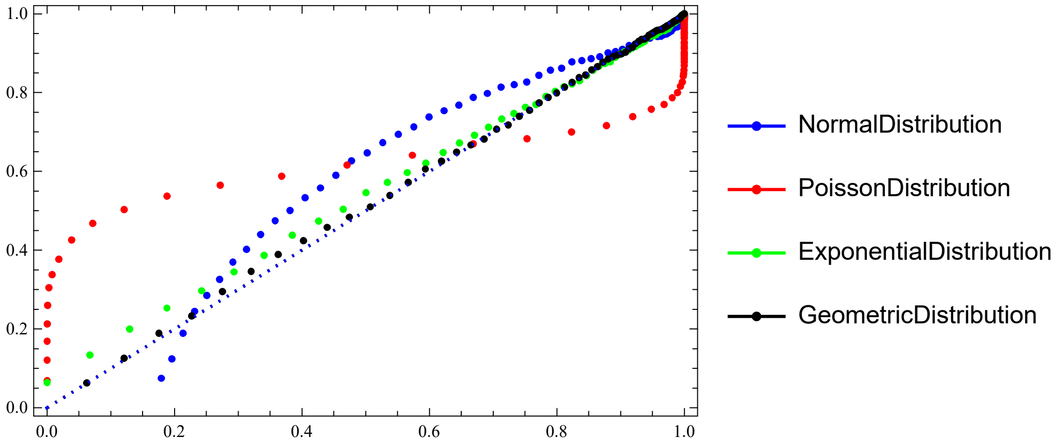
Bestäm fördelningsmodell

För hitta den rätta fördelningsmodellen för varje fördelning användes inbyggda funktionen EstimatedDistribution, som tog emot en lista av punkter (vår generade slumpvisa punkter)och en distributionsmodell.
Svaret från EstimatedDistribution lades in i ProbilityPlot funktionen med alla tagna punkter från randomNumber för att jämföra punkterna med fördelningsmodellen.

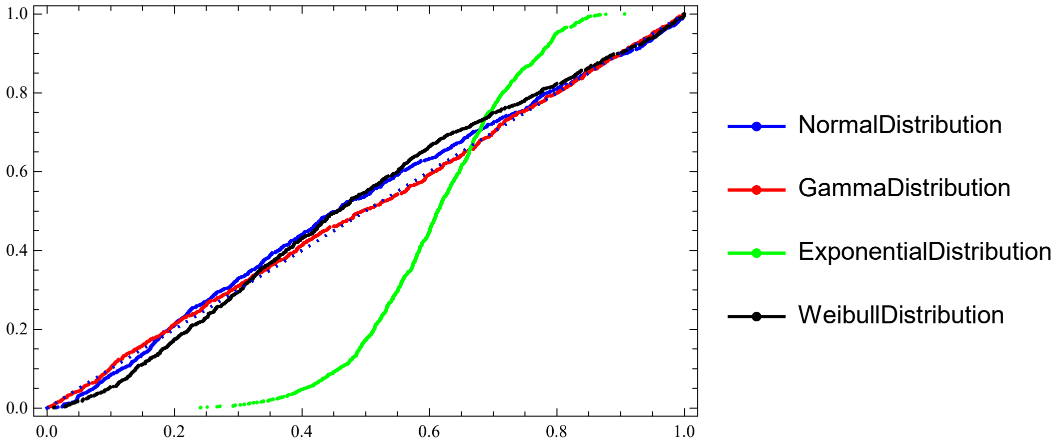
Vi hade tju modeller o välja mellan för varje fördelning, några fungerade att plotta ut medan andra kunde vi inte ens få ut eftersom vi fick felmeddelanden när vi försökte räkna ut dem.
De sju modellerna var normal-, gamma-, exponential-, weibull, geometrisk-, poisson- och binomialfördelning.



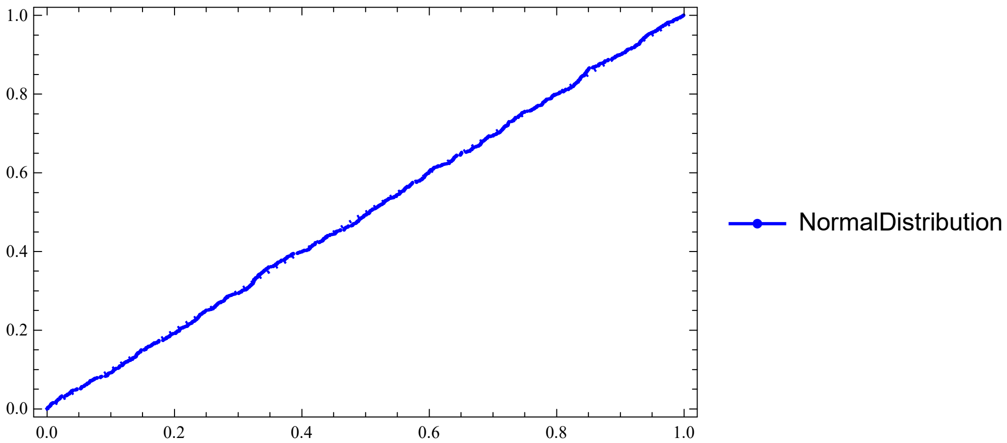
Vi kunde plotta ut fyra fördelningsmodeller för den första fördelningen. Man kan så att tre av dem är ganska lika varandra men den som är närmast räta linjen är gamma fördelningen.



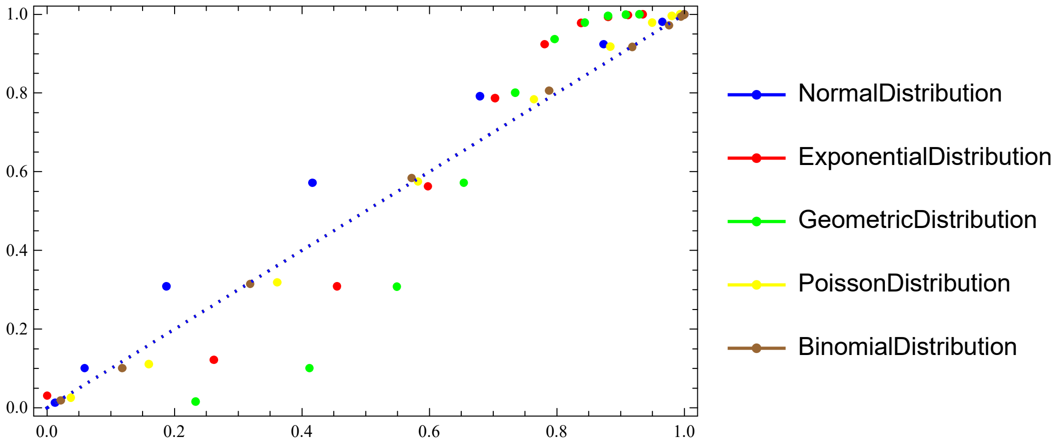
För den andra fördelningen kunde också få ut fyra modeller. Man kan se att den bästa fördelningsmodellen är gemotriska fördelningen.



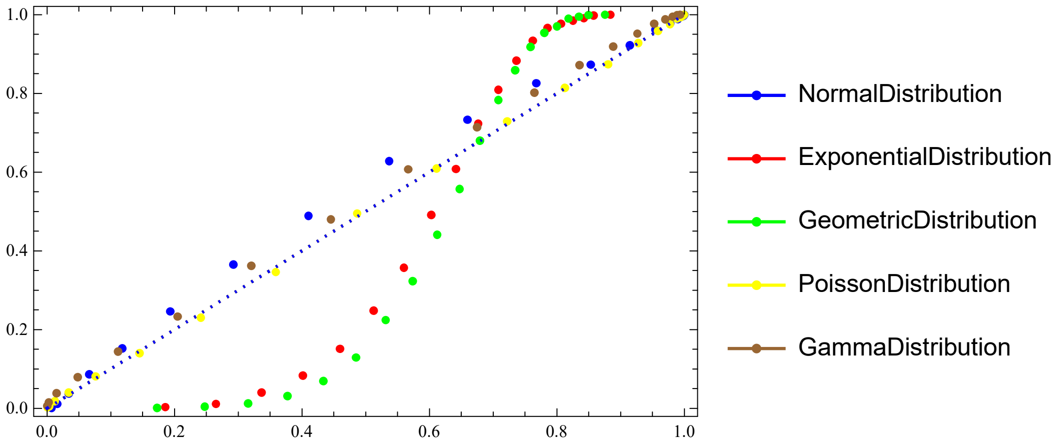
Gamma fördelningen passar bäst för den tredje fördelningen i randomNumber funktionen



Vi kunde bara få ut normal fördelningen för den fjärde fördelningen och vi kan se att modellen passar bra med fördelningen.



Man kan se att binomal fördelningen passar bäst för den femte fördelningen men vi kunde inte välja modellen eftersom i uppgift 6 (konfidensintervall) så fick vi felmeddelenade ibland när vi försökte ränka med binomial fördelningsmodellen.
Därför tog vi poisson fördelningen den bästa modell till femte fördelningen.



I grafen kan man se att poisson fördelningen är närmast räta linjen. Med det kan vi konstantera att den bästa modellen för sjätte fördelningen är poisson fördelning.

Beräkna konfidensintervaller

För att räkna ut konfidensintervaller för fördelningarna användes metoder tagna från `Hypothesis Testing Package`.

Vi har skapat en metod som får ut $\{\alpha, \beta\}$ genom `EstimatedDistribution` och använder sedan CLT (MeanCI metoden) för att få ut intervaller för α och β.
Detta görs om flera gånger tills CLT ger oss intervaller som inte är längre än dess medelvärde.

Kod