機率與統計簡介

機率與統計的關係(Relation between probability and statistic)

sample由某個機率分佈產生，而model由sample找出來

> Sample -----> Model: 統計(Statistic)
>
> Sample <----- Model: 機率(Probability)

基礎定理(Probability rules)

> 1. 互斥事件的加法(Addition rule for mutually exclusive events)：
>
> $$P(A\ or\ B) = P(A) + P(B)$$
>
> 2. 獨立事件的乘法(Multiplication rule for independent events)：
>
> $$P(A\ and\ B) = P(A) \cdot P(B)$$
>
> 3. 通用加法(General addition rule)：
>
> $$P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$$
>
> 4. 通用乘法(General multiplication rule)：
>
> $$P(A\ and\ B) = P(A)P(B|A)\ or\ P(B)P(A|B)$$

貝氏理論(Bayes theory)

- Let $A_{i=1\sim n}$ be mutually exclusive and exhaustive event.
- Then for any other event B(Law of total probability):

> $$P(A_k|B) = \frac{P(A_k\ and\ B)}{P(B)} = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

期望值與變異數計算(Expected value and Variance)

Item	Formula
期望值(Expected value)	$E(X) = \sum_{i=1}^{k} x_i p(x_i)$
變異數(Variance)	$V(X) = E(X^2) - E(X)^2 = \sum [x_i^2 p(x_i)] - \mu^2$
共變異數(Covariance)	$Cov(X, Y) = E(XY) - E(X)E(Y)$
相關係數(Correlation coefficient)	$Corr(X, Y) = \frac{Cov(X, Y)}{S.D(X) \cdot S.D(Y)}$

常見的機率分佈(Probability distibutions)

機率分佈大分類可分為

> 1. 離散型(Discrete)
> 2. 連續型(Continuous)

離散型(Discrete)

1. Bernoulli

   > 結果只有兩種(0 or 1，正面或反面)可能的實驗，例如投擲硬幣。

2. Geometric

   > 在伯努利實驗中，第一次成功所需要的實驗次數。

3. Binomial

   > 進行伯努利實驗n次，成功x次的機率分佈。

4. Poisson

   > 當Binomial的參數n很大，但p很小時的機率分佈。

	(r.v;參數)	期望值E	變異數	p.m.f
Bernoulli	(x;p)	p	p(1-p)	$P(x=1)=p$
Geometric	(x;p)	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$G(X=x)=p(1-p)^{x-1}$
Binomial	(x;n,p)	np	$np(1-p)$	$b(x;n,p)=\binom{n}{x}p^x(1-p)^{n-x}$
Poisson	(x;$\lambda>0$)	λ	λ	$p(x;\lambda)=\frac{e^{-\lambda}\lambda^x}{x!}$

連續型(Continuous)

1. Uniform

> 在指定的範圍(a,b)間，每個變數x的機率密度都是相同的機率分佈模型。

2. Normal

> 以平均數和標準差為參數的對稱機率分佈模型。

3. Exponential

> 以λ為參數的連續機率分佈，通常發生在描述獨立事件變數發生的時間間隔。

4. Gamma

> 假設X1, X2, ... Xn 為連續發生事件的等候時間，且這n次等候時間為獨立的，那麼這n次等候時間之和Y (Y=X1+X2+...+Xn)服從伽瑪分布，即 Y～Gamma(α , β)，亦可記作Y～Gamma(α , λ)，其中α = n，而 β 與λ互為倒數關係，λ 表單位時間內事件的發生率。 指數分布為α = 1的伽瑪分布。

5. Chi-Square(將(α,β)設成$(\frac{k}{2},2)$)

> 若k個隨機變數Z_1、Z_2...Z_k是相互獨立，符合標準常態分布的隨機變數（數學期望為0、變異數為1），則隨機變數Z的平方和$X=\sum_{i=1}^{k} Z_i^2$ 被稱為服從自由度為 k 的卡方分布，記作$X \sim \chi^2(k)$。

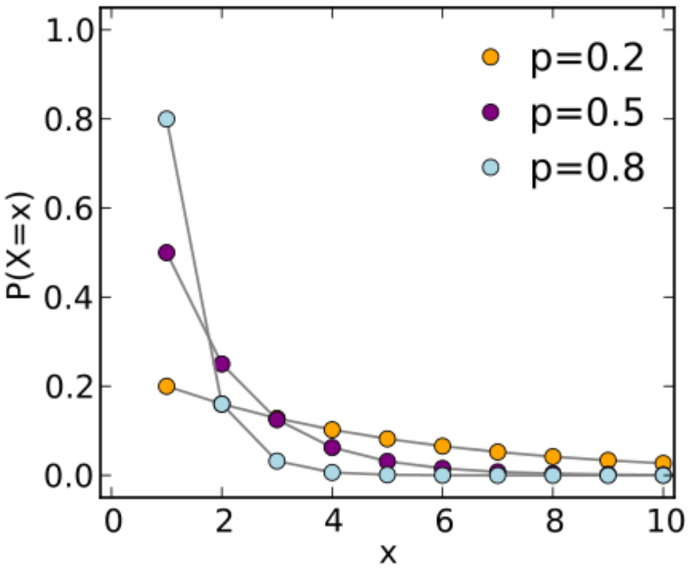
	(r.v;參數)	期望值E	變異數	p.d.f
Uniform	(x;a,b)	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$	$\frac{1}{b-a}$ for $a <= x <= b$
Normal	(x;μ,σ)	μ	σ^2	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
Exponential	(x,λ)	$\frac{1}{\lambda}=\beta$	$\frac{1}{\lambda^2}$	$f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x} & x\geq 0,\\ 0 & , x<0.\end{cases}$
Gamma	(x;α,β)	$\alpha\beta$	$\alpha\beta^2$	$f(x)=\frac{x^{(\alpha-1)}e^{\left(-\frac{1}{\beta}x\right)}}{\beta^\alpha\Gamma(\alpha)}, x>0$
Chi-Square	(x;k)	k	$2k$	$f_k(x)=\frac{\frac{1}{2}^{\frac{k}{2}}}{\Gamma(\frac{k}{2})}x^{\frac{k}{2}-1}e^{\frac{-x}{2}}$

機率分佈圖

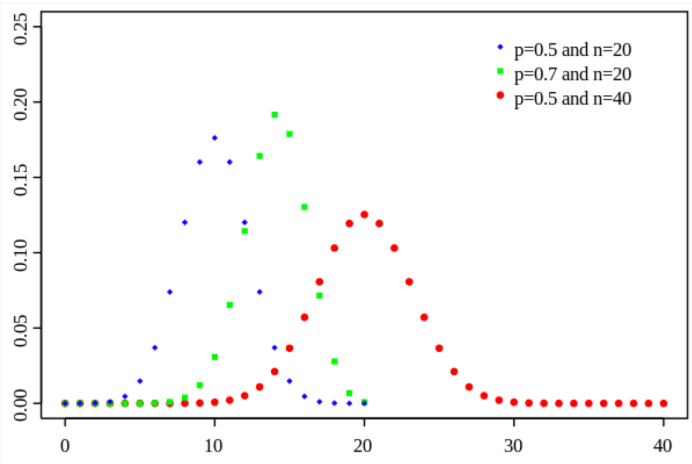
離散型

圖片來源:維基百科

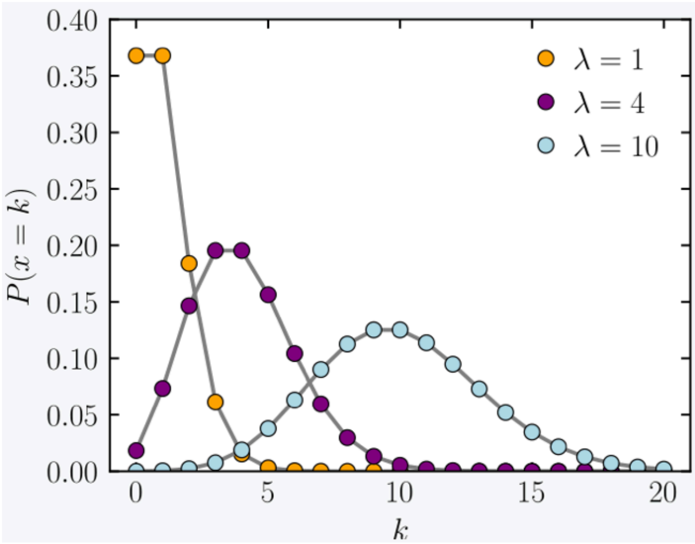
- Geometric distribution



- Binomial distribution



- Poisson distribution

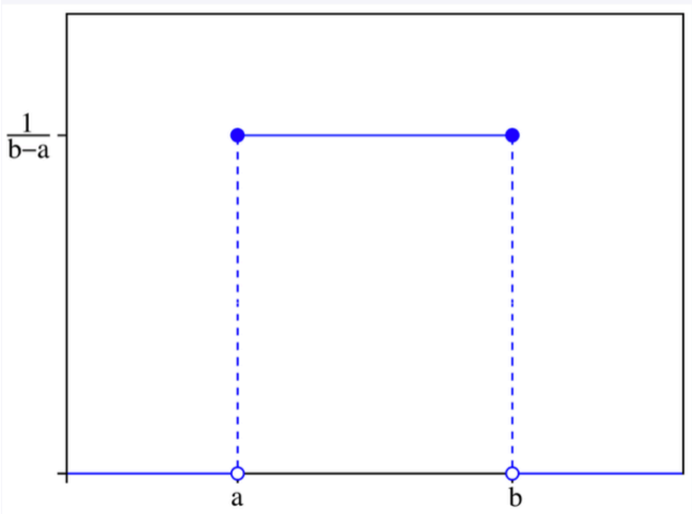


0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
$P(x = k)$
0
5
10
15
20
k
$\lambda = 1$
$\lambda = 4$
$\lambda = 10$

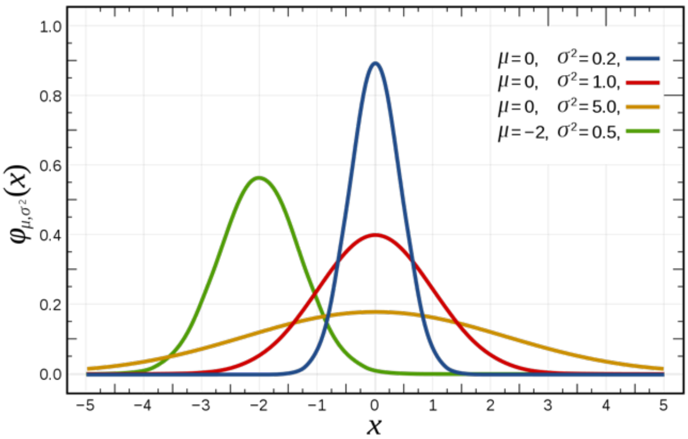
連續型

圖片來源:維基百科

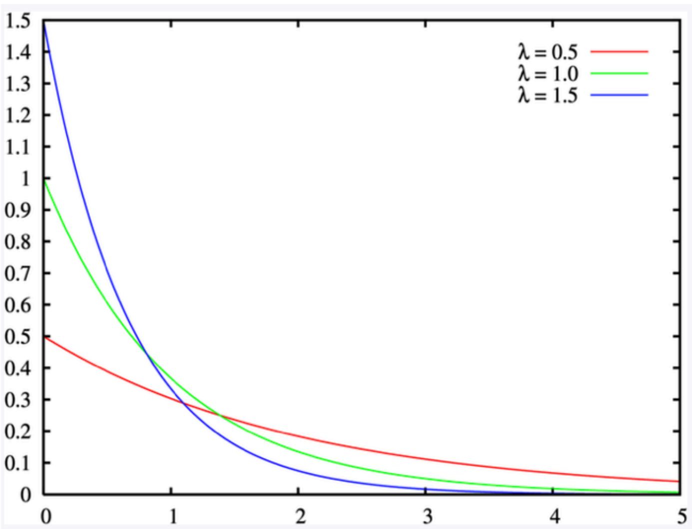
- Uniform distribution



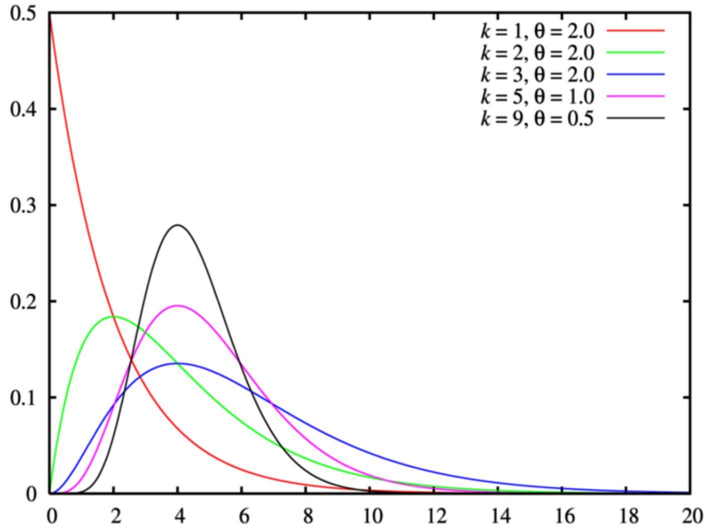
- Normal distribution



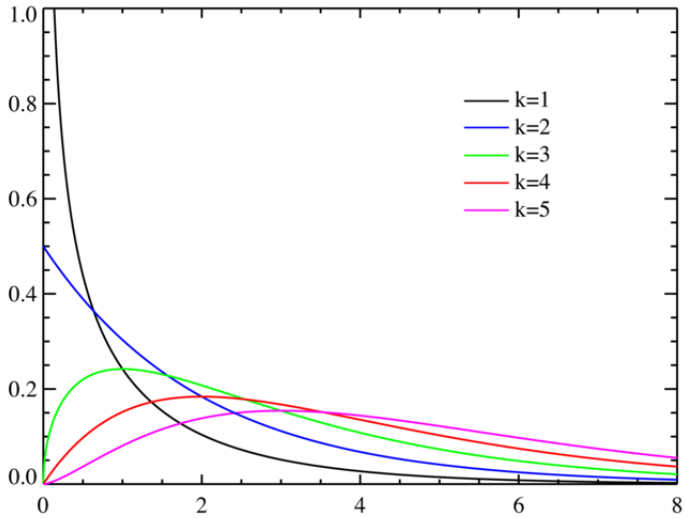
- Exponential distribution



- Gamma distribution



- chi-square distribution



In [ ]: