lesson 2

Conditional Prob. & Probability Distrib.

Todays topics

- Bayes' theorem
- Probability distributions

来嶋 秀治 (Shuji Kijima)

Dept. Informatics, ISEE

quick review of the last class & exercise

Probability Space

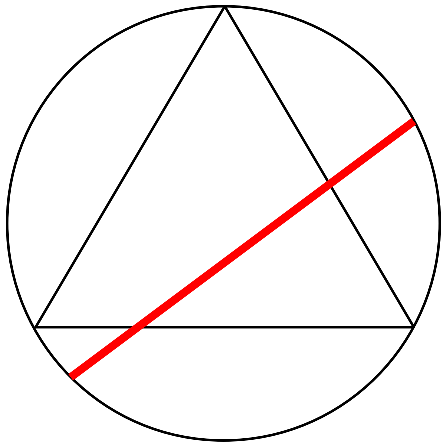
- ✓ Definitions
- ✓ Axiom
- ✓ Terminology

Ex 2. Bertrand paradox

- ✓ Consider an equilateral triangle inscribed in a circle.
- ✓ Suppose a chord of the circle is chosen at random.

Question

What is the probability that the chord is longer than a side of the triangle (=:x)?



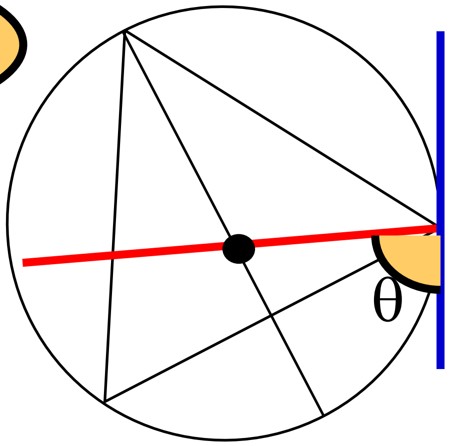
Ex 2. Bertrand paradox

- ✓ Consider an equilateral triangle inscribed in a circle.
- ✓ Suppose a chord of the circle is chosen at random.

Question

What is the probability that the chord is longer than a side of the triangle (=:x)?

What does "a chord of the circle is chosen at random" mean?



Definition: Probability Space

A probability space is defined by (Ω, F, P)

Ω: sample space(標本空間); a set of elementally events(標本点),

- an event(事象) is a subset of Ω.

F: σ-algebra ($\subseteq 2^{\Omega}$); a set of events.

P: probability measure(確率測度); a function $F \rightarrow R_{\geq 0}$,

- probability of an event.

Ex 2. Bertrand paradox

- ✓ Consider an equilateral triangle inscribed in a circle.
- ✓ Suppose a chord of the circle is chosen at random.

Question

What is the probability that the chord is longer than a side of the triangle (=:x)?

Answer 1: The "random radius" method:

✓ Choose a radius of the circle and a point on the radius.

the chord through this point and perpendicular to the radius.

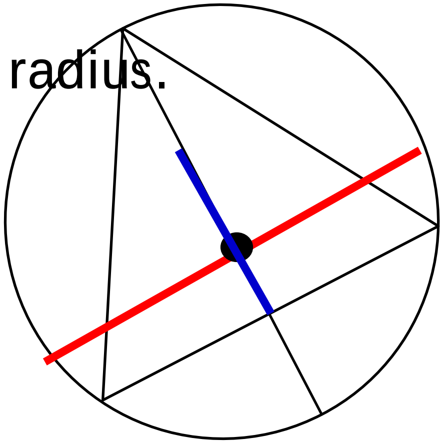
✓

⇒

$$\Omega_1 = \{d \in R \mid 0 \le d \le r\}$$

$$F_1 = 2^{\Omega_1}$$

$$P_1(d \le y) = \frac{y}{r}$$



Ex 2. Bertrand paradox

- ✓ Consider an equilateral triangle inscribed in a circle.
- ✓ Suppose a chord of the circle is chosen at random.

Question

What is the probability that the chord is longer than a side of the triangle (=:x)?

Answer 2: The "random endpoints" method::

✓ Choose two random points on the circumference of the circle and draw the chord joining them.

✓

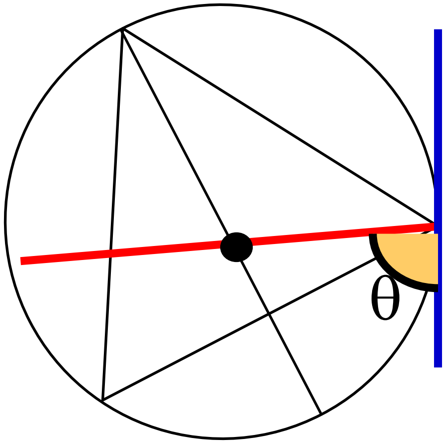
⇒

w.l.o.g one end point is (r,0)

$$\Omega_2 = \{\theta \in R \mid 0 \le \theta < \pi\}$$

$$F_2 = 2^{\Omega_2}$$

$$P_2(\theta \le y) = \frac{y}{\pi}$$



Ex 2. Bertrand paradox

- ✓ Consider an equilateral triangle inscribed in a circle.
- ✓ Suppose a chord of the circle is chosen at random.

Question

What is the probability that the chord is longer than a side of the triangle (=:x)?

Answer 3: The "random midpoint" method.

✓ Choose a point anywhere within the circle, and construct a chord with the chosen point as its midpoint.

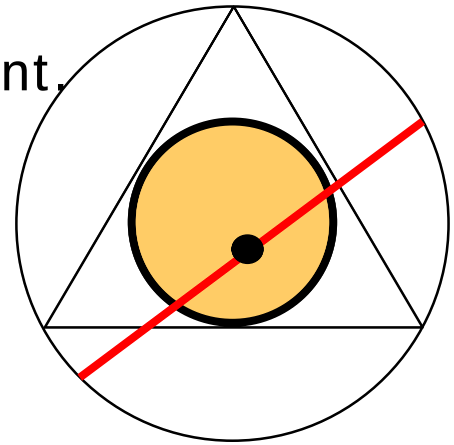
✓

$$\Omega_3 = \{(a, b) \in R^2 \mid a^2 + b^2 \le r^2\}$$

$$F_3 = 2^{\Omega_3}$$

$\Rightarrow$

$$P_3(a^2 + b^2 \le x^2) = \left(\frac{x}{r}\right)^2$$



Ex. 3. Boy or Girl

Question 1.

Desmond and Molly has two kids. One is a boy.

What is the probability that the other is a girl?

Today's topic 1

Conditional Probability

def.s

- ✓ joint probability
- ✓ conditional probability
- ✓ independence / mutually independence

thm.

- ✓ Bayes' theorem
- ✓ Answer for the Monty Hall Problem

Terminology

✓Def. 1. Joint probability; (同時確率 or 結合確率)

$$\Pr(A, B) = \Pr(A \cap B)$$

✓Def. 2. Conditional Probability (条件付き確率)

$$\Pr(A \mid B) = \frac{\Pr(A, B)}{\Pr(B)}$$

✓Def. 3. Events A and B are independent (独立)

see ex. 1.

$$\Pr(A, B) = \Pr(A)\Pr(B)$$

✓Events $A_1, A_2, \dots, A_k$ are mutually independent (相互に独立)

$$\Pr\left(\cap_{i=1}^{k} A_i\right) = \prod_{i=1}^{k} \Pr(A_i)$$

✓Events $A_1, A_2, \dots, A_k$ are pairwise independent (対ごとに独立)

$$\Pr\left(A_i, A_j\right) = \Pr(A_i)\Pr\left(A_j\right) \quad \text{for any distinct } i, j$$

Tossing coins (Independence)

Suppose two coins.

Head probability of coin A is 0.5.

Head probability of coin B is 0.5.

The probability of two heads

$$\Pr([\mathrm{H}],[\mathrm{H}]) = \Pr([\mathrm{H}])\Pr([\mathrm{H}]) = \square$$

Tossing coins (Independence)

Suppose two coins.

Head probability of coin A is 0.6.

Head probability of coin B is 0.7.

	H	T	Prob.
H	0.42	0.18	0.6
T	0.28	0.12	0.4
Prob.	0.7	0.3	

The probability of two heads

$$\Pr([H],[H]) = \Pr([H])\Pr([H]) = \square$$

Tossing coins (Dependence)

Two coins are made of magnets.

Head probability of coin A is 0.5.

Head probability of coin B is 0.5.

Tossing coins (Dependence)

Two coins are made of magnets.

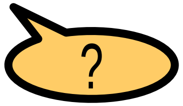
Head probability of coin A is 0.5.

Head probability of coin B is 0.5.

	H	T	Prob.
H	0.05	0.45	0.5
T	0.45	0.05	0.5
Prob.	0.5	0.5	

The probability of two heads

$$\Pr([\mathrm{H}],[\mathrm{H}]) = \Pr([\mathrm{H}])\,\Pr([\mathrm{H}])$$



Independence test

	Good (early healing)	No good	Total
Med.	28	22	50
Placebo	13	37	50
Total	41	59	100

$$\Pr([\text{med.}],[\text{good}]) = \Pr([\text{med}])\,\Pr(\text{good})$$

Bayes’ theorem

Thm. (Bayes; ベイズ)

$$\Pr(A \mid B) = \frac{\Pr(B \mid A)\Pr(A)}{\Pr(B)}$$

Bayes' theorem (general)

Thm. (Bayes; ベイズ)

$A_1, \ldots, A_k$ are mutually exclusive, and $\cup_{i=1}^{k} A_i = \Omega$.

$$\Pr(A_i \mid B) = \frac{\Pr(B \mid A_i)\Pr(A_i)}{\sum_{j=1}^{k} \Pr(B \mid A_j)\Pr(A_j)}$$

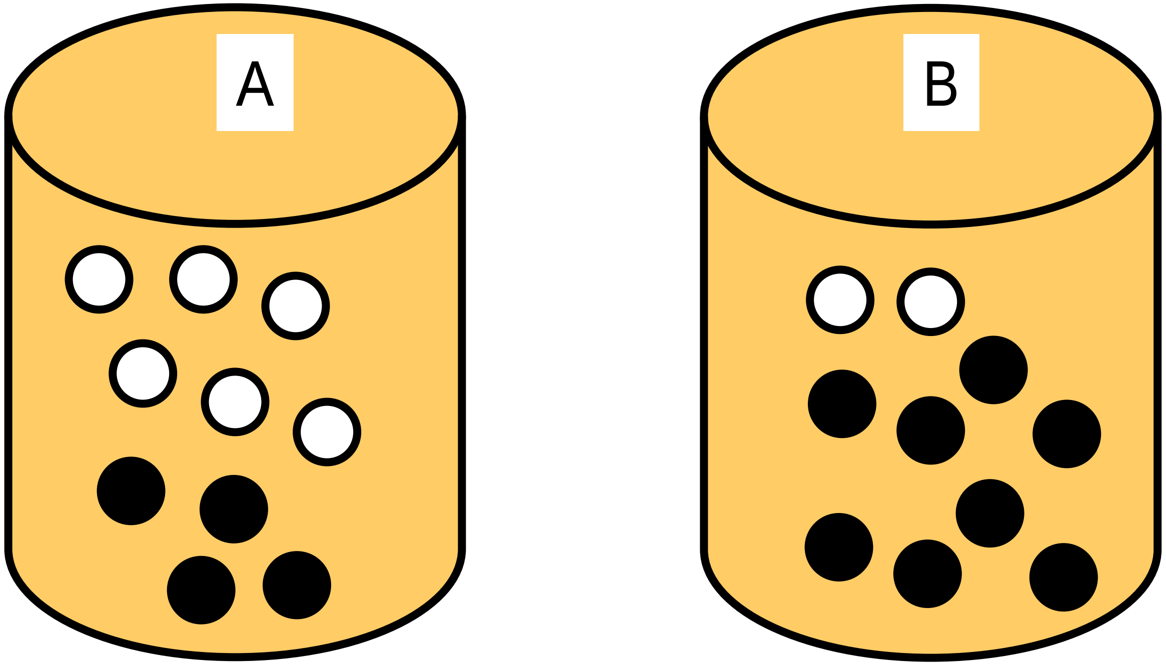
Prop.

$A_1, \ldots, A_k$ are mutually exclusive, and $\cup_{i=1}^{k} A_i = \Omega$.

$$\Pr(B) = \sum_{i=1}^{k} \Pr(A_i, B)$$

(the right hand side) is called marginal distribution.

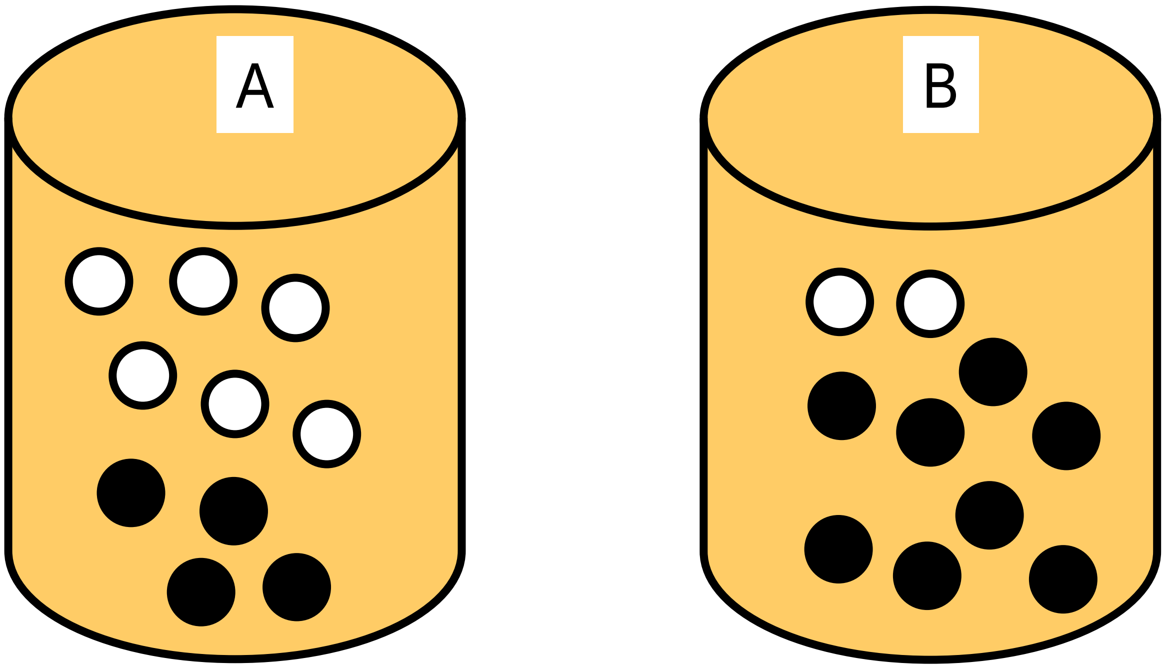
Conditional Probability



Conditional Probability

$\Pr(\circ \mid A) =$

Bayes' probability



Bayes' probability

$\Pr(A|\circ) =$

Ex. 3. Boy or Girl

Question 1.

Desmond and Molly has two kids. One is a boy.

What is the probability that the other is a girl?

Ex. 3. Boy or Girl

Question 1.

Desmond and Molly has two kids. One is a boy.

What is the probability that the other is a girl?

	Elder	Younger	Prob.
Case 1	Boy	Boy	1/4
Case 2	Boy	Girl	1/4
Case 3	Girl	Boy	1/4
Case 4	Girl	Girl	1/4

$$\Pr[\,\mathrm{G} \mid \mathrm{B}\,] = \frac{\Pr[\mathrm{B},\mathrm{G}]}{\Pr[B]} = \frac{2/4}{3/4} = \frac{2}{3}$$

Ex 1. Monty Hall problem --- ask Marilyn

- ✓ You are given the choice of three doors:
  - ➢ Behind on door is a car; behind the others goats.
- ✓ You pick a door, say A.
- ✓ The host (Monty), who knows what's behind the doors, opens another door, say C, which he knows has a goat.
- ✓ He then says to you, "Do you want to pick door B?"

Question

Is it to your advantage to switch your choice?

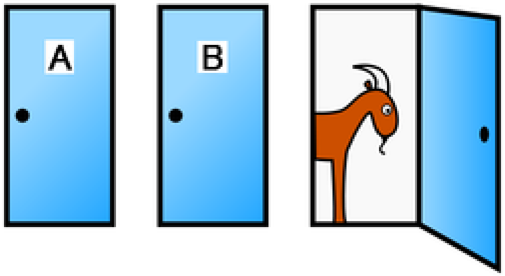


図: wikipedia”モンティーホール問題”より

Ex 1. Monty Hall problem --- ask Marilyn

- ✓ You are given the choice of three doors:
  - ➢ Behind on door is a car; behind the others goats.
- ✓ You pick a door, say A.
- ✓ The host (Monty), who knows what's behind the doors, opens another door, say C, which he knows has a goat.
- ✓ He then says to you, "Do you want to pick door B?"

$$P(B^* \mid C) = \frac{P(C \mid B^*)P(B^*)}{P(C \mid A^*)P(A^*) + P(C \mid B^*)P(B^*)} = \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{2}{3}$$

cf. ex. 3.

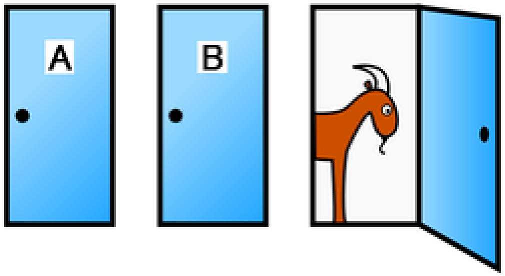


図: wikipedia”モンティーホール問題”より

Today's topic 2

Discrete Probability & Expectation

def.s

- ✓ discrete random variable
- ✓ discrete distribution
- ✓ expectation / conditional expectation

thm.

- ✓ Linearity of the expectation
- ✓ Coupon collector

"variable" vs "random variable"

Ex. 1. Set Ω

✓ $\Omega=\{1,2,3,4,5,6\}$,

Let x be a member of Set Ω.

Observation

➢ $x \in \Omega$

Def. random variable

Ex. 1. die (Ω, F, P)

- ✓ Ω={1,2,3,4,5,6},
- ✓ $F=2^{\Omega}$
- ✓ P(A)=|A|/6 (for any $A \subseteq \Omega$).

Let X denote the "cast" of (Ω, F, P)

Observation

- X ∈ Ω (∈ F in fact)
- P(X is odd) = 1/2
- P(X < 5) = 2/3 etc.

called random variable.
(usually denoted by CAPITALS)

Note

random variable may not be a member of F.

- e.g., Let Y:= square of cast

where, there is a map from F. (see regime)

terminology

X is called "random variable (確率変数)"

note Ξ may not be Ω (cf. ex. 6)

Discrete distribution (離散分布)

distribution on countable set $\Xi \subseteq R$ such that

$\Sigma_{x \in \Xi}$ Pr(X=x) = 1 holds

important concept in continuous distr. (next week)

✓ Probability function (確率関数)

f(x) = Pr(X=x)

✓ (cumulative) distribution function ((累積)分布関数)

F(x) = Pr(X $\leq$ x)

Expectation of random variable X (確率変数の期待値)

E(X) = $\Sigma_{x \in \Xi}$ x Pr(X=x)

Conditional expectation (相互条件付き期待値)

E(X | Y) = $\Sigma_{x \in \Xi}$ x Pr(X=x | Y=y)

(univariate) discrete distributions

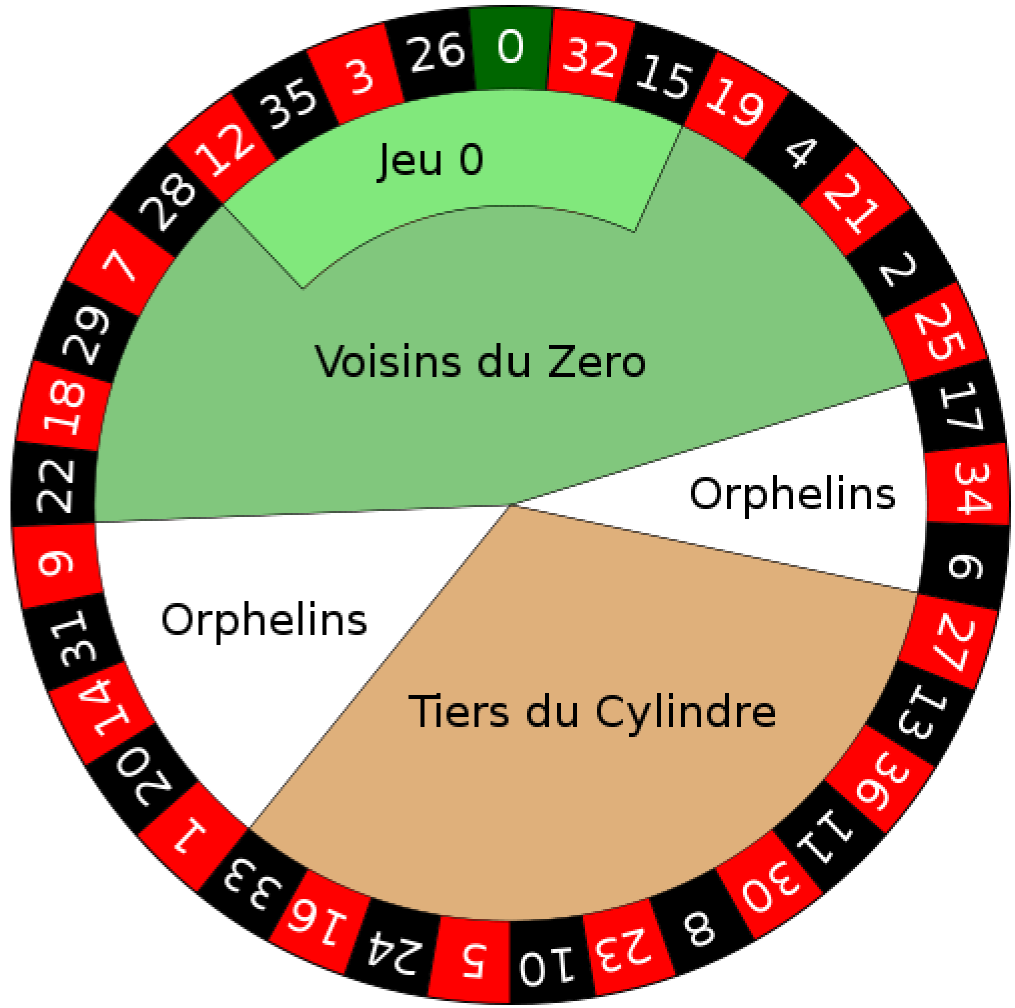
- ✓ uniform dist. (離散一様分布)
- ✓ Bernoulli dist. (ベルヌーイ分布; 2点分布)
- ✓ binomial dist. (2項分布)
- ✓ geometric dist. (幾何分布)
- ✓ Poisson dist. (ポアソン分布)

discrete uniform (離散一様分布)

$\Omega=\{1,2,\ldots,n\}$

$P(X = i) = 1/n$

roulette



$\Omega = \{0,1,\ldots,36\}$

$F = 2^{\Omega}$

$P(x) = 1/37 \quad (x \in \Omega)$

Bernoulli (ベルヌーイ分布, 2点分布) B(1;p)

$\Omega=\{0,1\}$

$P(X = 1) = p$

$P(X = 0) = 1-p$

An experiment outputting a random variable according to Bernoulli dist. is said **Bernoulli trial (ベルヌーイ試行)**.

(biased) coin tossing

head (X=1)

tail (X=0)

binomial dist. (2項分布) B(n;p)

Ω={0,1,2,…,n}

$$\Pr[X = k] = \binom{n}{k} p^k (1-p)^{n-k}$$

Let $X_1, X_2, \ldots, X_n$ be outputs of Bernoulli trial (B(1;p)), i.i.d.

Let $X = X_1 + X_2 + \ldots + X_n$,

- meaning that the total number of heads.

X is according to a binomial distribution B(n;p)

geometric dist. (幾何分布) Ge(p)

$\Omega=\{0,1,2,\ldots\}$

$P(X = k) = (1-p)^k\, p$

Repeat Bernoulli trials B(1;p) i.i.d., until head.

Let K denote the number of tail before head,

then K is according to a geometric distribution Ge(p).

Remember coupon collector.

Poisson dist. (ポアソン分布) Po(λ) (λ>0)

Ω={0,1,2,...}

$$\Pr[X = z] = e^{-\lambda} \frac{\lambda^z}{z!}$$

Let's consider the probability of rare events,
the expected number of occurrences is λ in a unit time.
Let X be the number of occurrences,
then X is known to be according to the Poisson distr. Po(λ).

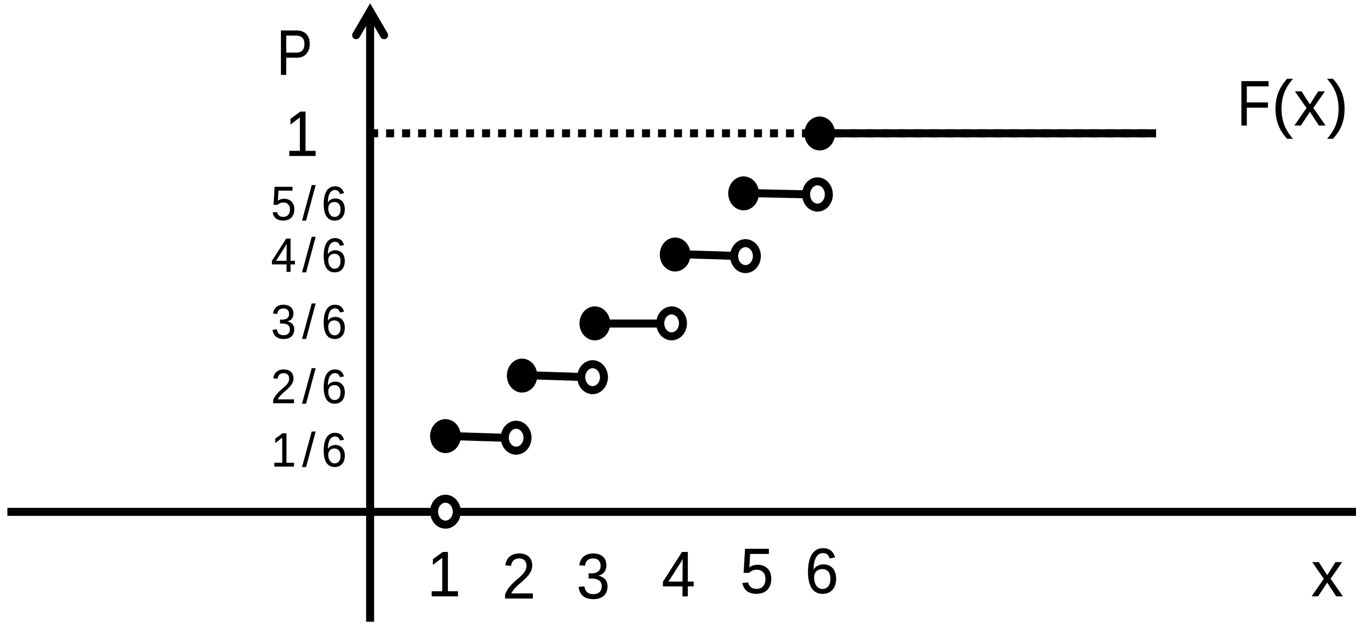
More precisely, repeat Bernoulli trials B(1;p) i.i.d. with p << 1.
Let λ=np, then it is known that B(n;p) ≈ Po(λ).
⇒ today's Exercise 2. Poisson distr. appears later today.

Discrete distr.: (distr. on a countable set $\Omega \subset R$)

- ✓ $\Sigma_{x \in \Omega}$ Pr(X=x) = 1 holds.
- ✓ probability function (確率関数)

  f(x) = P(X=x)
- ✓ (cumulative) distribution function ((累積)分布関数)

  F(X) = P(X ≤ x)

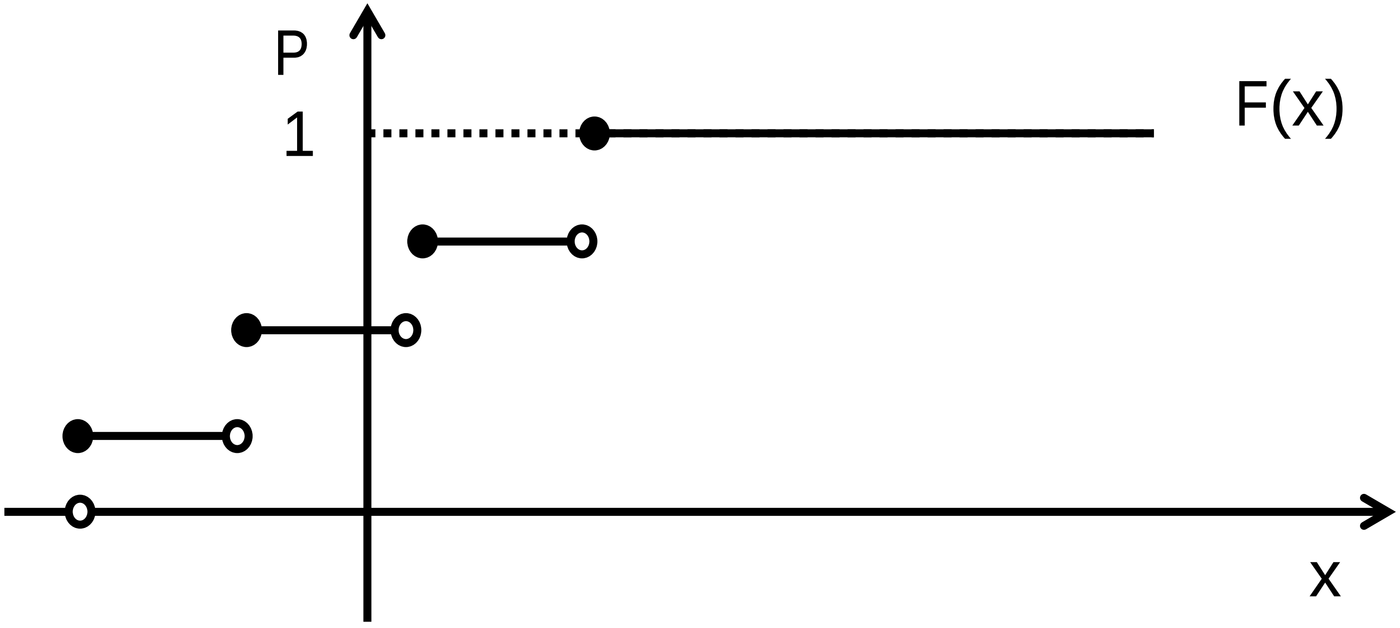


Discrete distr.: (distr. on a countable set $\Omega \subset R$)

- ✓ $\Sigma_{x \in \Omega}$ Pr(X=x) = 1 holds.
- ✓ probability function (確率関数)

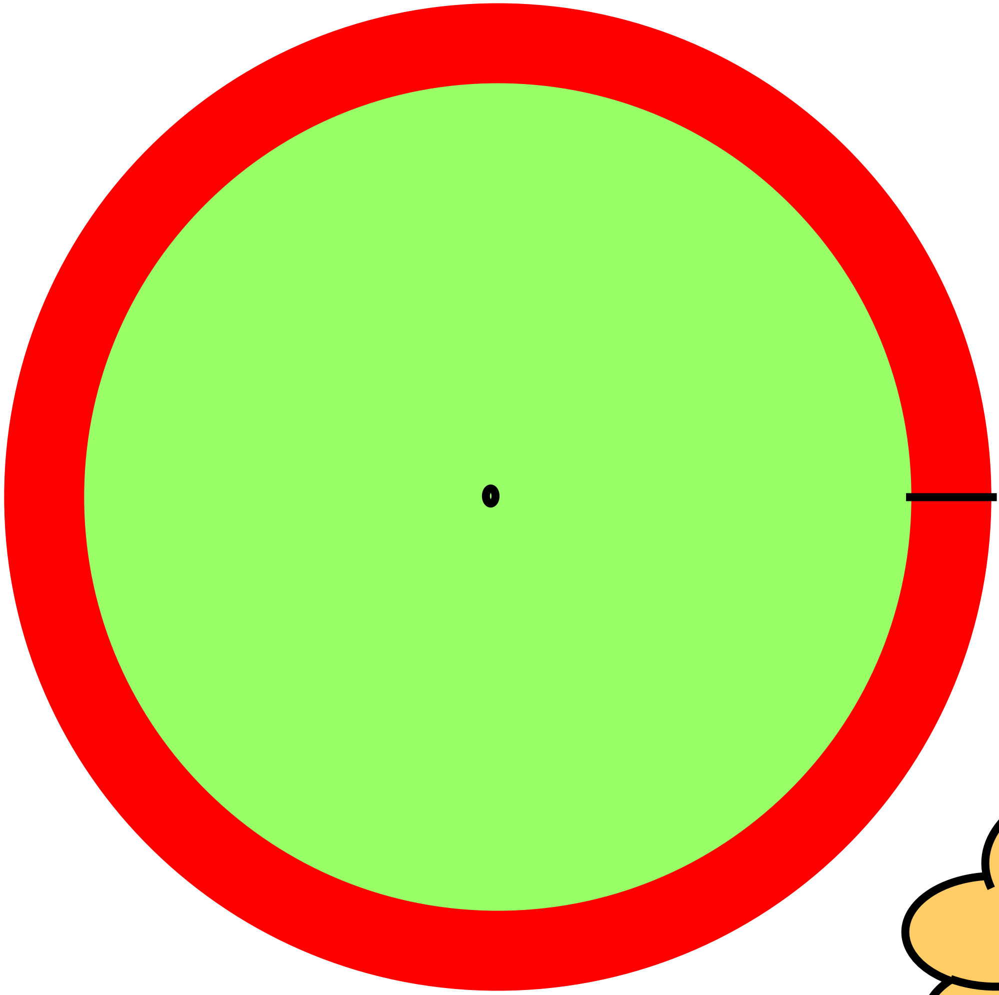
  f(x) = P(X=x)
- ✓ (cumulative) distribution function ((累積)分布関数)

  F(X) = P(X $\leq$ x)



(univariate) continuous distributions

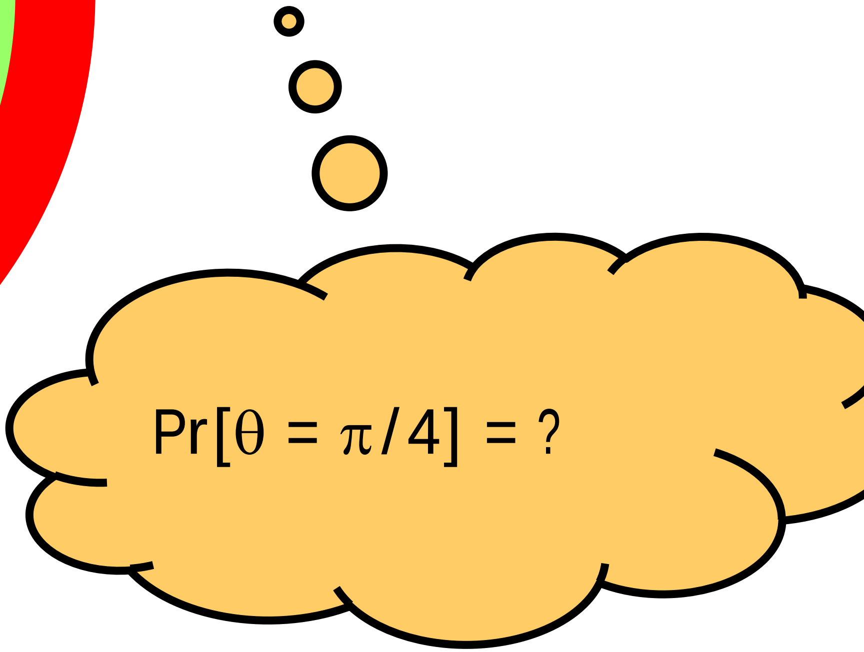
Continuous roulette



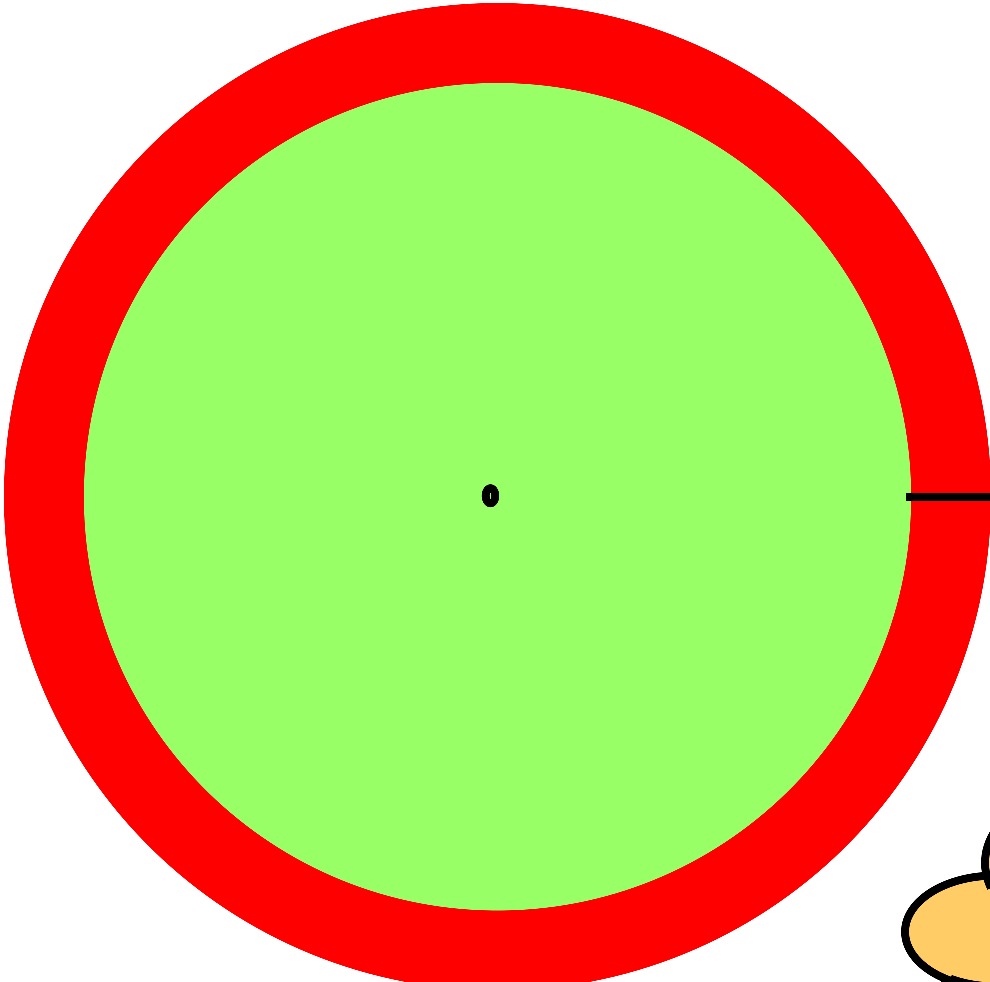
$\Omega = \{\theta \mid 0 \leq \theta < 2\pi\}$

$F = 2^{\Omega}$

$P(x) = ?$ $(x \in \Omega)$



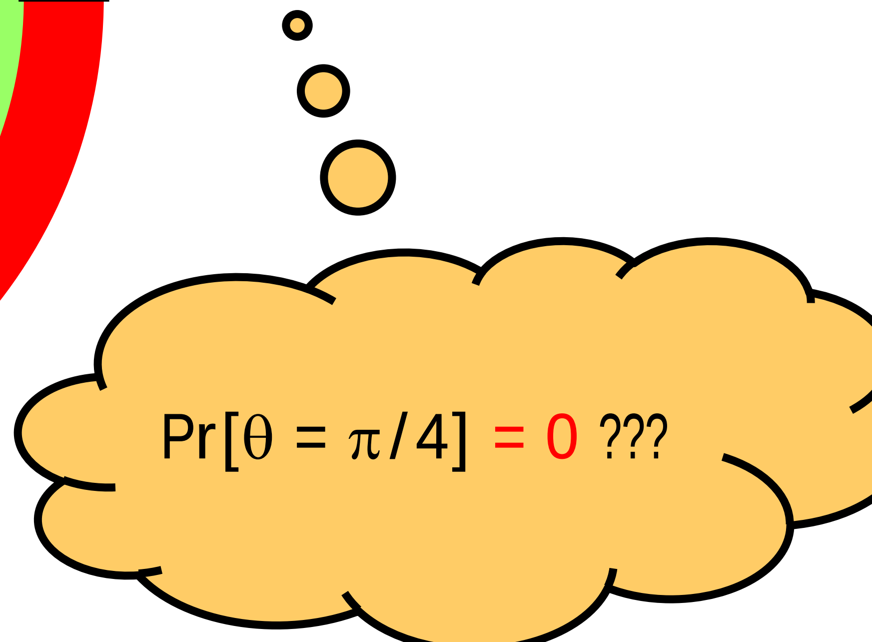
Continuous roulette



$\Omega = \{\theta \mid 0 \leq \theta < 2\pi\}$

$F = 2^{\Omega}$

$P(x) =$ **?** $(x \in \Omega)$



(continuous) uniform distr.

$\Omega=(0,2\pi]$

$P(X = i) = ???$

$P(X \leq x) = 1/x$

cumulative distribution function

seems appropriate.

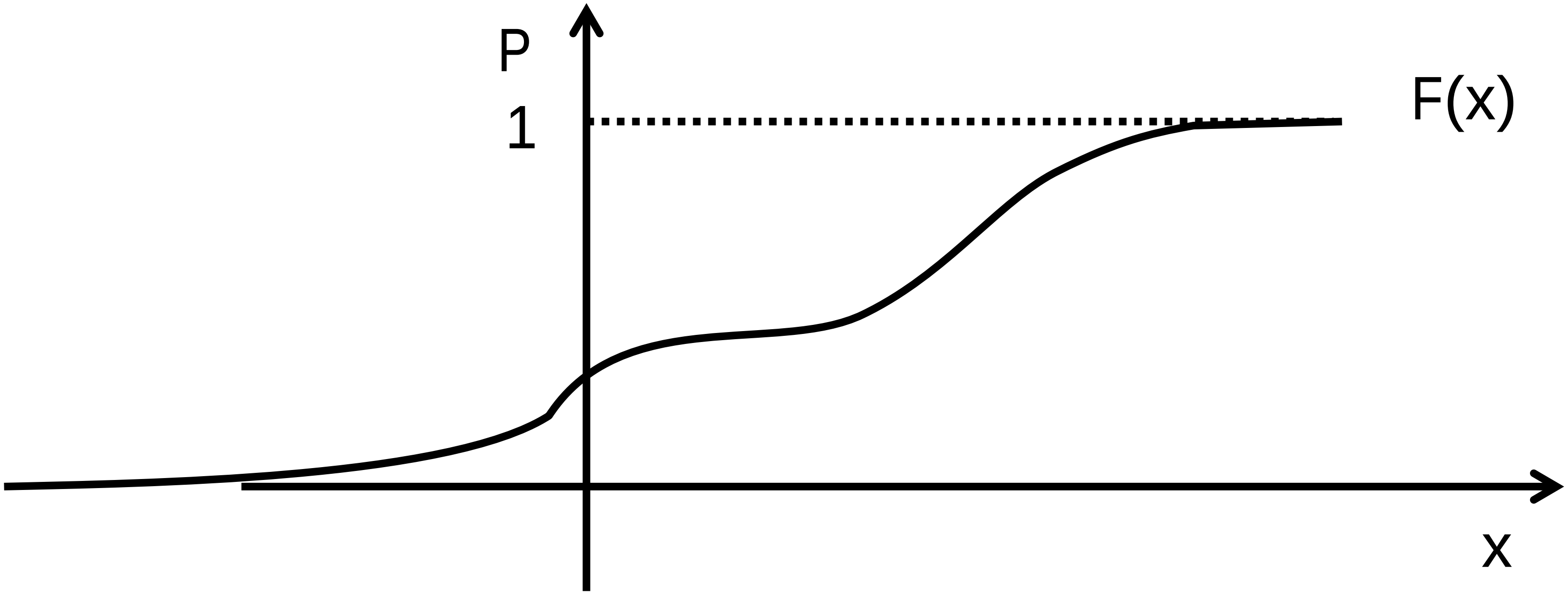
continuous distr. (distr. on uncountable set $\Omega \subset R$)

- ✓ probability density function (確率密度関数)

  $f(x) = dF(x)/dx$

- ✓ (cumulative) distribution function ((累積)分布関数)

  $F(X) = P(X \leq x)$ differentiable (continuous)



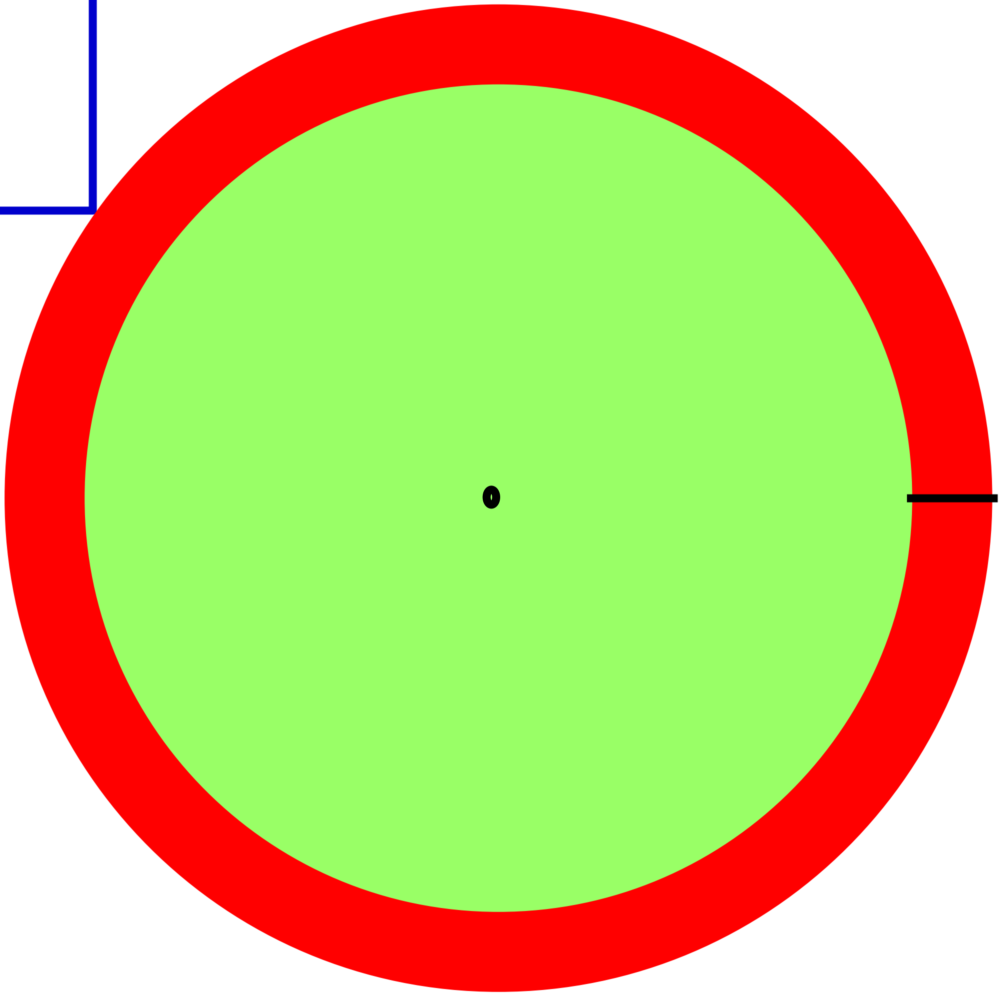
Uniform ditr. (一様分布) U(a,b)

Ω=(a,b]

$$f(x) = \frac{1}{b-a} \quad (a \le x \le b)$$

$$F(x) = \frac{x-a}{b-a} \quad (a \le x \le b)$$

continuous roulette



$\Omega = (0, 2\pi]$

$F = 2^{\Omega}$

$f(x) = 1/2\pi \quad (x \in \Omega)$

Normal distr. (正規分布) N(μ, σ²)

Ω=(−∞, +∞)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \quad (-\infty < x < \infty)$$

Exponential distr. (指数分布) Ex(λ) (λ>0)

Ω=(0, +∞)

$$f(x) = \lambda \mathrm{e}^{-\lambda x} \quad (x \geq 0)$$

Gamma distr. (ガンマ分布) G(α,ν) (α>0, ν>0)

Ω=(0, +∞)

$$f(x) = \frac{1}{\Gamma(\nu)} \alpha^{\nu} x^{\nu-1} \mathrm{e}^{-\alpha x} \quad (x \geq 0)$$

where

$$\Gamma(\nu) = \int_{-\infty}^{\infty} t^{\nu-1} \mathrm{e}^{-t} \mathrm{d}t$$

remark that

$\Gamma(1) = 1$

$\Gamma(\nu) = (\nu - 1)\Gamma(\nu - 1)$

$\Gamma(\nu) = (\nu - 1)! \quad (\nu = 1, 2, \dots)$

Exponential distr., Gamma distr., and Poisson distr.

Proposition 1

$G(\lambda,1) = Ex(\lambda)$

consider the case that ν_1 and ν_2 are integers.

Proposition 2

Suppose $X_1 \sim G(\alpha,\nu_1)$ and $X_2 \sim G(\alpha,\nu_2)$ are independent,

then $X := X_1+X_2 \sim G(\alpha,\nu_1+\nu_2)$

Proposition 3

Suppose $X_1, X_2, \ldots \sim Ex(\lambda)$, i.i.d.,

then $N := \max\{ n \mid X_1+\ldots+X_n \le 1\} \sim Po(\lambda)$

Exponential distr., Gamma distr., and Poisson distr.

Proposition 2

Suppose $X_1 \sim G(\alpha,\nu_1)$ and $X_2 \sim G(\alpha,\nu_2)$ are independent, then $X:=X_1+X_2 \sim G(\alpha,\nu_1+\nu_2)$

consider the case that ν_1 and ν_2 are integers.

$$
\begin{aligned}
f_X(x) &= \int_0^x f_{X_1}(t) f_{X_2}(x-t)\mathrm{d}t \\
&= \int_0^x \frac{1}{\Gamma(\nu_1)} \alpha^{\nu_1} t^{\nu_1-1} \mathrm{e}^{-\alpha t} \frac{1}{\Gamma(\nu_2)} \alpha^{\nu_2} (x-t)^{\nu_2-1} \mathrm{e}^{-\alpha(x-t)} \mathrm{d}t \\
&= \int_0^x \frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} \cdot \frac{1}{\Gamma(\nu_1+\nu_2)} \alpha^{\nu_1+\nu_2} \mathrm{e}^{-\alpha x} t^{\nu_1-1} (x-t)^{\nu_2-1}\, \mathrm{d}t \\
&= \frac{1}{\Gamma(\nu_1+\nu_2)} \alpha^{\nu_1+\nu_2} \mathrm{e}^{-\alpha x} \int_0^x \frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} t^{\nu_1-1} (x-t)^{\nu_2-1} \mathrm{d}t
\end{aligned}
$$

Exponential distr., Gamma distr., and Poisson distr.

Proposition 2

Suppose $X_1 \sim G(\alpha,\nu_1)$ and $X_2 \sim G(\alpha,\nu_2)$ are independent,
then $X := X_1 + X_2 \sim G(\alpha,\nu_1+\nu_2)$

consider the case that ν_1 and ν_2 are integers.

let $s = t/x$, i.e., $t = xs$, then $\mathrm{d}t = x\mathrm{d}s$ and

$$\int_0^x \frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} t^{\nu_1-1}(x-t)^{\nu_2-1}\mathrm{d}t$$

$$= \int_0^1 \frac{(\nu_1+\nu_2)!}{\nu_1!\,\nu_2!}(xs)^{\nu_1-1}(x-xs)^{\nu_2-1}x\mathrm{d}s$$

$$= x^{\nu_1+\nu_2-1}\int_0^1 \frac{\Gamma(\nu_1+\nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} s^{\nu_1-1}(1-s)^{\nu_2-1}\mathrm{d}s$$

$$= x^{\nu_1+\nu_2-1}$$

Exponential distr., Gamma distr., and Poisson distr.

Proposition 3

Suppose $X_1, X_2, \ldots \sim \mathrm{Ex}(\lambda)$, i.i.d.,

then $N \coloneqq \max\{n \mid X_1 + \cdots + X_n \leq 1\} \sim \mathrm{Po}(\lambda)$

$$f_N(n) = \int_0^1 \Pr[X_1 + \cdots + X_n \leq 1, X_{n+1} > 1 - t \mid X_1 + \cdots + X_n = t] f_{\mathrm{Ga}(\lambda,n)}(t) dt$$

$$= \int_0^1 \Pr[\, X_{n+1} > 1 - t \mid X_1 + \cdots X_n = t\,] f_{\mathrm{Ga}(\lambda,n)}(t) \mathrm{d}t$$

$$= \int_0^1 \Pr[X_{n+1} > 1 - t] f_{\mathrm{Ga}(\lambda,n)}(t) \mathrm{d}t$$

$$= \int_0^1 \left(1 - F_{X_{n+1}}(t)\right) \frac{1}{(n-1)!} \lambda^n t^{n-1} \mathrm{e}^{-\lambda t} \mathrm{d}t$$

$$= \int_0^1 e^{-\lambda(1-t)} \frac{1}{(n-1)!} \lambda^n t^{n-1} \mathrm{e}^{-\lambda t} \mathrm{d}t$$

$$= \mathrm{e}^{-\lambda} \int_0^1 \lambda \frac{(\lambda t)^{n-1}}{(n-1)!} \mathrm{e}^{-\lambda t} dt = \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!}$$