Lesson 3

Expectation, Variance, Moment…

Todays topics

- **expectation**,
- **Markov's inequality**
- **variance**, **covariance**, **moment**
- **Chebyshev's inequality**

来嶋 秀治 (Shuji Kijima)
Dept. Informatics,
Graduate School of ISEE

Today's topic 1

Probability Distributions

<u>def.s</u>

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

<u>thm.</u>

- ✓ Linearity of the expectation
- ✓ Coupon collector

“variable” vs “random variable”

Ex. 1. Set Ω

✓ Ω={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)

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

Let X denote the "cast" of (Ω, F, P)

Observation

- ➢ $X \in \Omega$ ($\in F$ 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

✓ Probability function (確率関数)

$f(x) = \Pr(X=x)$

important concept in continuous distr. (next week)

✓ (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)

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

(univariate) discrete distributions

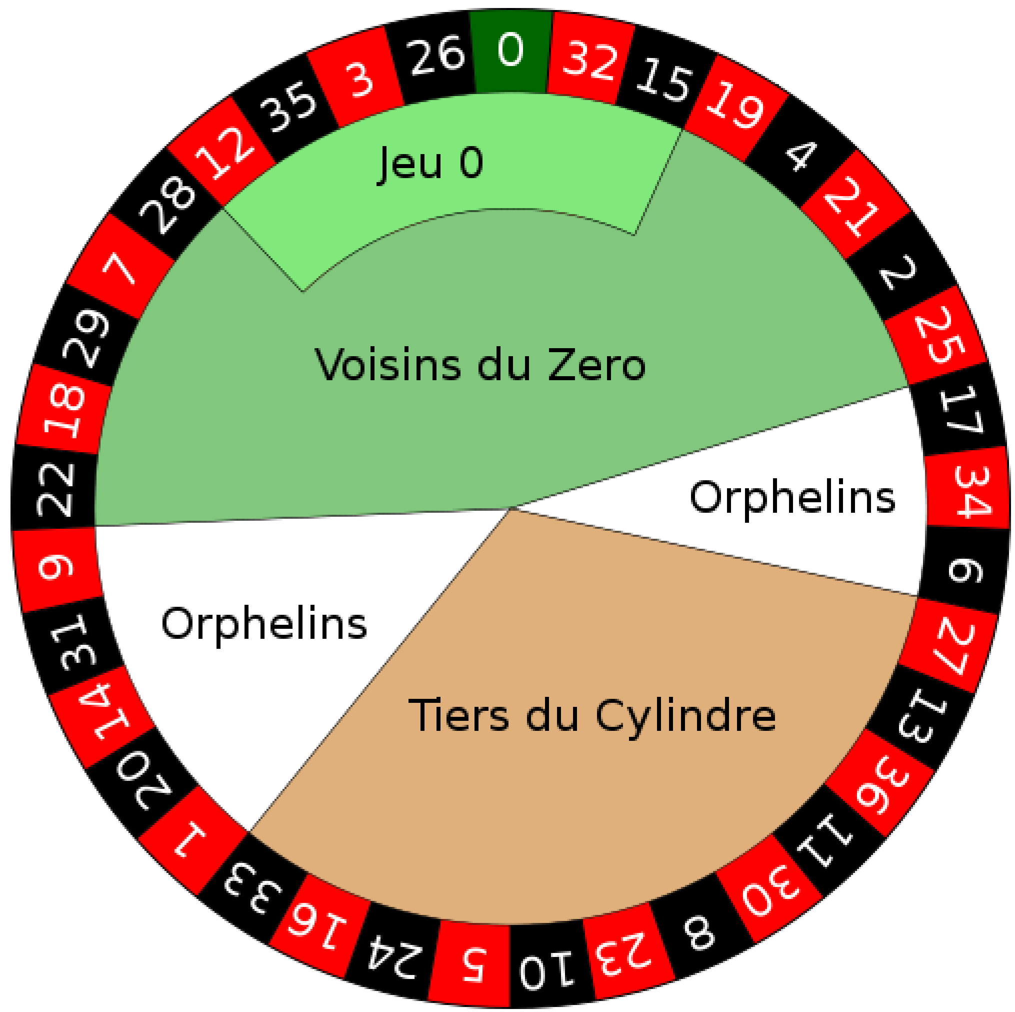
- ✓ uniform dist. (離散一様分布)
- ✓ Bernoulli dist. (ベルヌーイ分布; 2点分布)
- ✓ binomial dist. (2項分布)
- ✓ geometric dist. (幾何分布)
- ✓ Poisson dist. (ポアソン分布)

discrete uniform (離散一様分布)

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

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

roulette



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

$F = 2^{\Omega}$

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

Bernoulli (ベルヌーイ分布, 2点分布) B(1;p)

Ω={0,1}

P(X = 1) = p

P(X = 0) = 1-p

(biased) coin tossing

head (X=1)

tail (X=0)

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

binomial dist. (2項分布) B(n;p)

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

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

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)

Ω={0,1,2,...}

$$f(k) = \Pr[X = k] = (1-p)^k p$$

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,...}

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

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)

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

  $f(x) = \Pr[X = x]$
- ✓ (cumulative) distribution function ((累積)分布関数)

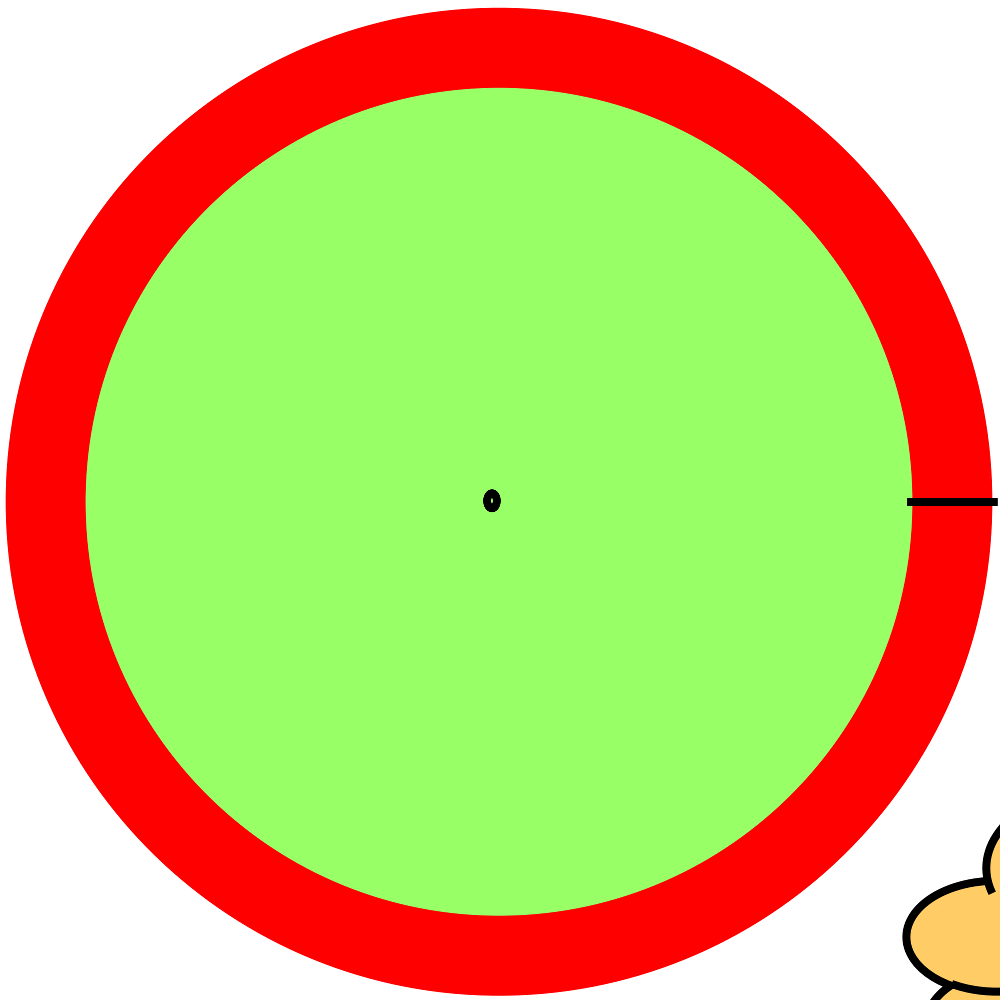
  $F(x) = \Pr[X \leq x]$

Discrete Distribution Function $F: \Omega \rightarrow \mathrm{R}_{\geq 0}$

1. $F(-\infty) = 0,\ F(+\infty) = 1$
2. Monotone non-decreasing (単調非減少)
3. Right continuous (右連続)

(univariate) continuous distributions

Continuous roulette



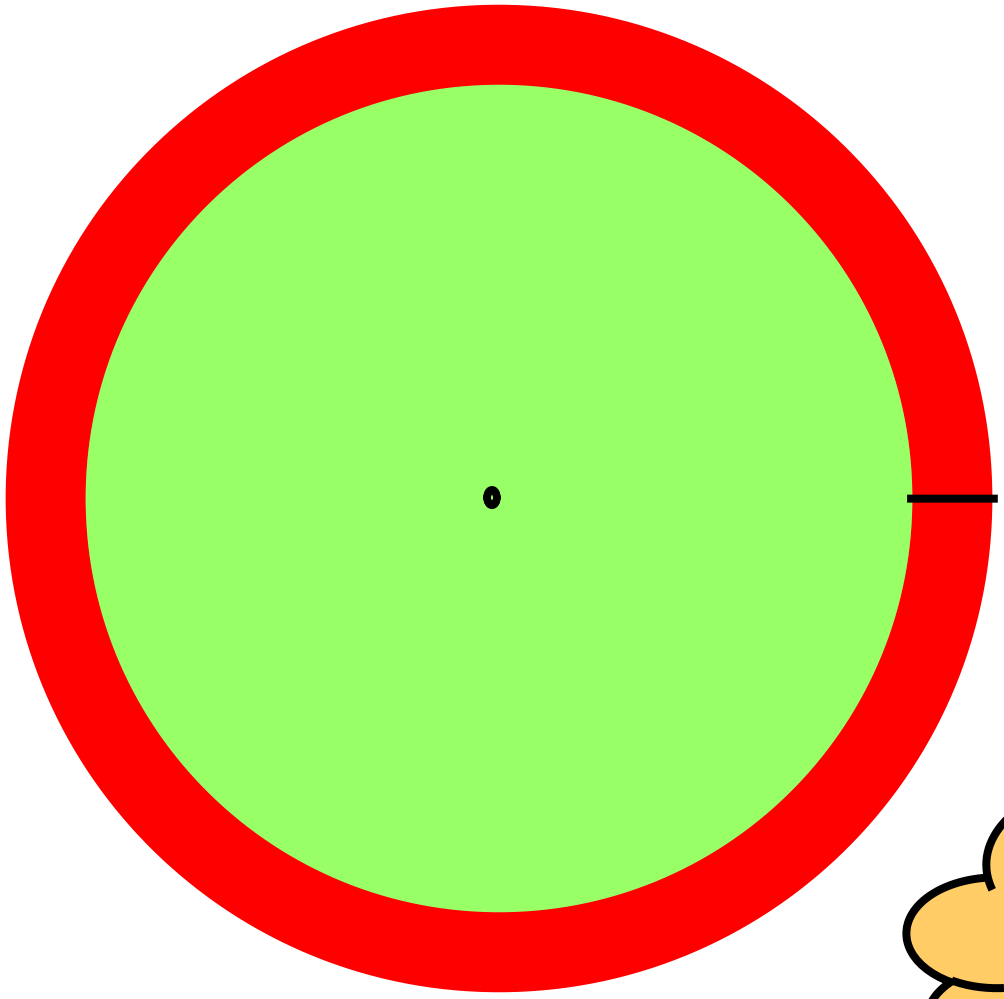
$\Omega = \{\theta \mid 0 \leq \theta < 2\pi\}$

$F = 2^{\Omega}$

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

$Pr[\theta = \pi/4] = ?$

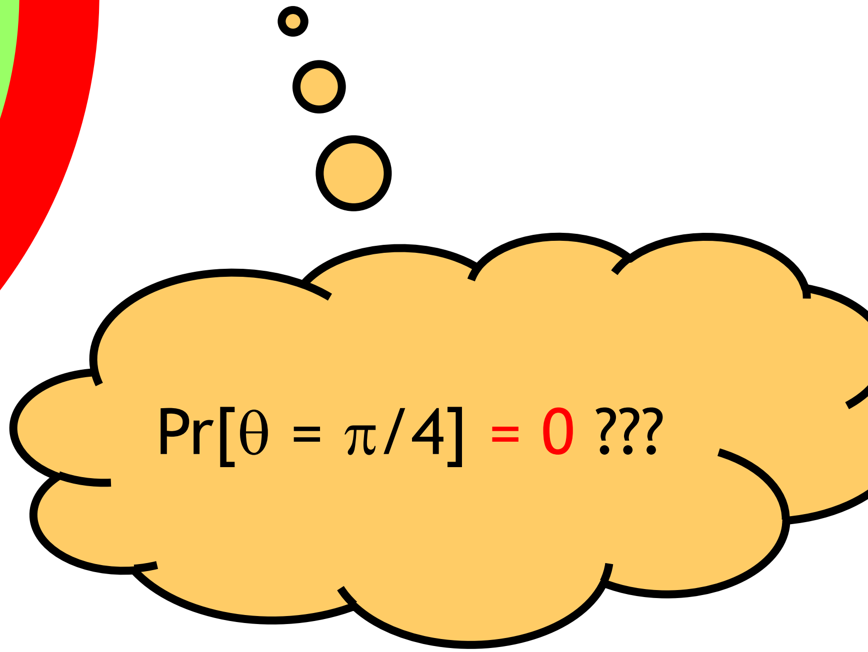
Continuous roulette



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

$F = 2^{\Omega}$

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



(continuous) uniform distr.

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

$P(X = x) = ???$

$P(X \leq x) = x/2\pi$

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)

Continuous Distribution Function $F: \mathrm{R} \to \mathrm{R}_{\geq 0}$

1. $F(-\infty) = 0,\ F(+\infty) = 1$
2. Monotone non-decreasing (単調非減少)
3. Differentiable* (微分可能)

   *in the effective domain.

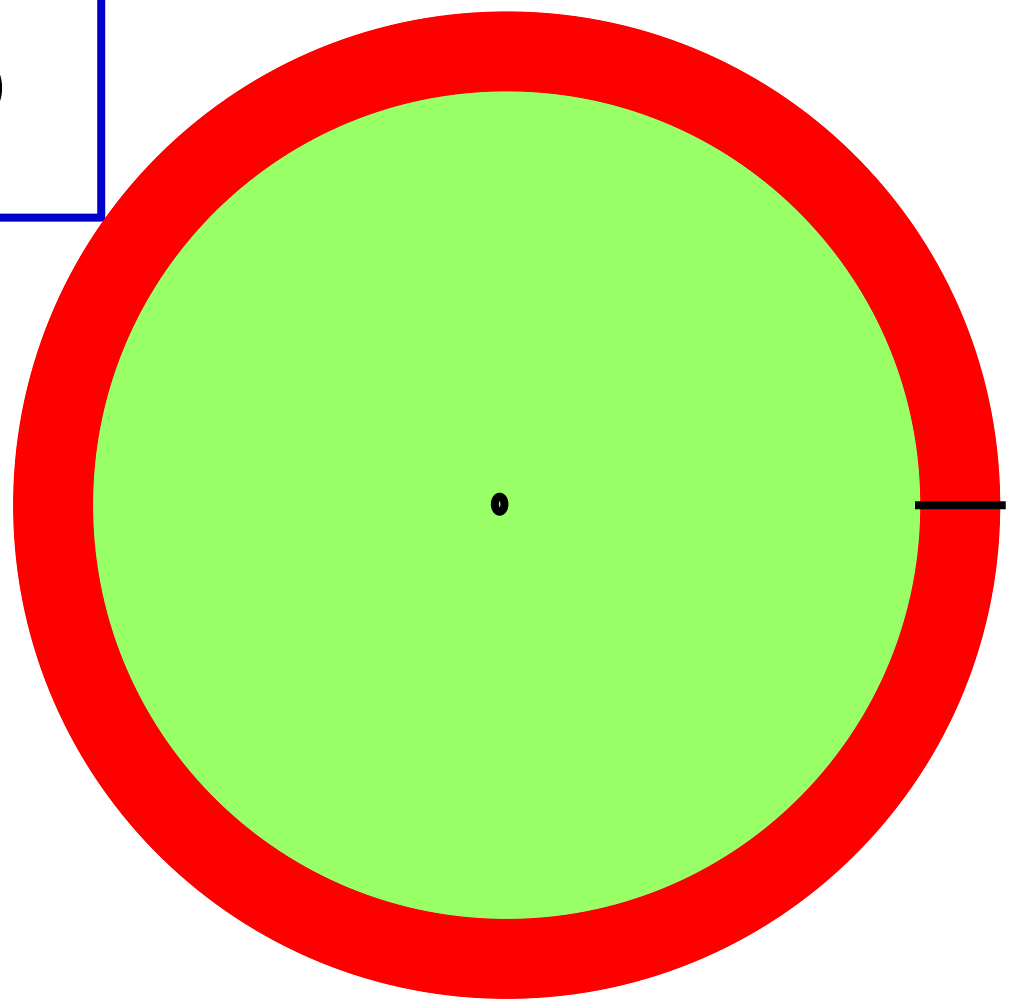
Uniform ditr. (一様分布) U(a,b)

Ω=(a,b]

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

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

continuous roulette



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

$\mathcal{F} = 2^{\Omega}$

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

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

Normal distr. (正規分布) N(μ, σ^2)

$\Omega=(-\infty,\ +\infty)$

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

Exponential distr. (指数分布) Ex(λ) (λ>0)

$\Omega=(0, +\infty)$

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

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

$\Omega=(0, +\infty)$

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

where

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

remark that

$\Gamma(1) = 1$

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

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

Some Distributions

Discrete distributions

(1) Bernoulli B(1,p)

(2) Binomial B(n,p)

- #heads during tossing n coins.

(3) Geometric Ge(p)

- # tails before a head.

(4) Poisson Po(λ)

Continuous distributions

(1) Uniform U(a,b)

(2) Exponential Ex(α)

(3) Normal N(μ,σ^2)

(4) Beta Be(α,β)

(5) Gamma G(θ,k)

i.i.d.

Distribution of random variables X and Y of $(\Omega, \mathcal{F}, P)$.

<u>Ex1. two dice.</u>

$\Omega = \{(1,1),(1,2),\ldots,(6,5),(6,6)\}$

X = sum of casts

Y = product of casts

<u>例2. poker</u>

choose five cards,

X = # of A's

Y = # of spades

Joint Distribution (同時分布; 結合分布)

✓Joint distribution

$$F(x, y) \coloneqq \Pr[(X \le x), (Y \le y)]$$

(pdf: $f(x, y) \coloneqq \frac{\partial^2}{\partial x \partial y} F(x, y)$)

cf. multivariate

cf. multivariate distribution

multivariate discrete distribution

distr. fnc. : $F(x,y) \coloneqq \Pr[(X,Y) \leq (x,y)]\,(= \Pr[(X \leq x),(Y \leq y)])$

pmf: $f(x,y) \coloneqq \Pr[(X,Y) = (x,y)]\,(= \Pr[(X = x),(Y = y)])$

multivariate continuous distribution

distr. fnc. : $F(x,y) \coloneqq \Pr[(X,Y) \leq (x,y)]\,(= \Pr[(X \leq x),(Y \leq y)])$

pdf: $f(x,y) \coloneqq \frac{\partial^2}{\partial x \partial y} F(x,y)$

terminology 2

✓X and Y are independent (独立)

$$F_{XY}(x,y) = F_X(x)F_Y(y)$$

prop. X,Y independent $\Rightarrow f_{XY}(x,y) = f_X(x)f_Y(y)$

✓X,Y are identically distributed (同一分布に従う)

f(x) = f(y)

✓ X,Y are independent and identically distributed

(i.i.d.;独立同一分布)

Prop.

Prop.

$$F_{XY}(x,y) = F_X(x)F_Y(y) \Rightarrow f_{XY}(x,y) = f_X(x)f_Y(y).$$

Proof.

$$f(x,y) := \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)$$

$$= \frac{\partial^2}{\partial x \partial y}\left(F_X(x)F_Y(y)\right)$$

$$= \frac{\partial}{\partial x}\left(\left(\frac{\partial}{\partial y}F_X(x)\right)F_Y(y)\right) + \frac{\partial}{\partial x}\left(F_X(x)\left(\frac{\partial}{\partial y}F_Y(y)\right)\right)$$

$$= 0 + \frac{\partial}{\partial x}\left(F_X(x)f_Y(y)\right)$$

$$= \frac{\partial}{\partial x}F_X(x)f_Y(y) + F_X(x)\frac{\partial}{\partial x}f_Y(y)$$

$$= f_X(x)f_Y(y)$$

Final question. Coupon collect

- •ビックリマンシール
- •ポケモンカード

✓ The are n kinds of coupons.

✓ How many coupons do you need to draw, in expectation, before having drawn each coupon at least once ?

Today’s topic 2

Expectation, variance, moment

Expectation

- ✓ Expectation (期待値) of a discrete random variable X is defined by

$$\mathrm{E}[X] = \sum_{x \in \Omega} x \cdot f(x)$$

only when the right hand side is converged absolutely (絶対収束),

i.e., $\sum_{x \in \Omega} |x \cdot f(x)| < \infty$ holds.

If it is not the case, we say "expectation does not exist."

- ✓ Expectation (期待値) of a continuous random variable X is defined by

$$\mathrm{E}[X] = \int_{-\infty}^{+\infty} x \cdot f(x) \mathrm{d}x \, .$$

Compute expectations of distributions

*Ex 2.

Discrete

(*i) Bernoulli distribution $B(1,p)$.

(*ii) Binomial distribution $B(n,p)$.

(iii) Geometric distribution $Ge(p)$.

(iv) Poisson distribution $Po(\lambda)$.

Continuous

(v) Exponential distribution $Ex(\alpha)$.

(vi) Normal distribution $N(\mu,\sigma^2)$.

Ex. Expectation of Geom. distr.

Thm.

The expectation of X~B(n,p) is np

proof

$$\sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} k \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} k \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{n!}{(k-1)!\,(n-k)!} p^k (1-p)^{n-k}$$

$$= \sum_{k=1}^{n} np \frac{(n-1)!}{(k-1)!\,(n-k)!} p^{k-1} (1-p)^{n-k}$$

$$= np \sum_{k'=0}^{n-1} \binom{n-1}{k'} p^{k'} (1-p)^{n-1-k'}$$

$$= np$$

Ex. Expectation of Geom. distr.

Thm.

The expectation of $X \sim Ge(p)$ is $(1-p)/p$

proof

$$E[X] = 0p + 1(1-p)p + 2(1-p)^2p + 3(1-p)^3p + \dots$$

$$(1-p)E[X] = (1-p)p + 1(1-p)^2p + 2(1-p)^3p + \dots$$

$$pE[X] = (1-p)p + (1-p)^2p + (1-p)^3p + \dots$$

$$= (1-p)p/(1-(1-p)) = 1-p.$$

Thus, $E[X] = (1-p)/p$.

Properties of Expectations

Thm.

For an arbitrary constant c,

$$E[c] = c$$

$$E[cX] = cE[X]$$

$$E[X+c] = E[X]+c$$

Linearity of expectations (discrete random variables)

Thm. (linearity of expectation; 期待値の線形性)

$$\mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i)$$

proof.

$$\mathrm{E}[X+Y]$$

$$= \sum_x \sum_y (x+y) \Pr([X=x] \cap [Y=y])$$

$$= \sum_x \sum_y (x+y) f(x,y)$$

$$= \sum_x \sum_y x f(x,y) + \sum_x \sum_y y f(x,y)$$

$$= \sum_x x \sum_y f(x,y) + \sum_y y \sum_x f(x,y)$$

$$= \sum_x x f(x) + \sum_y y f(y)$$

$$= \mathrm{E}[X] + \mathrm{E}[Y]$$

Linearity of expectations (continuous random variables)

Thm. (linearity of expectation; 期待値の線形性)

$$\mathrm{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{E}(X_i)$$

proof.

$$\begin{aligned}
&\mathrm{E}[X+Y] \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)f(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)\mathrm{d}x\mathrm{d}y + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty}x\left(\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y\right)\mathrm{d}x + \int_{-\infty}^{+\infty}y\left(\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x\right)\mathrm{d}y \\
&= \int_{-\infty}^{+\infty}xf(x)\mathrm{d}x + \int_{-\infty}^{+\infty}yf(y)\mathrm{d}y \\
&= \mathrm{E}[X] + \mathrm{E}[Y]
\end{aligned}$$

Application of linearity of expectation

Thm.

The expectation of $X \sim B(n;p)$ is np

proof

Suppose $X_1,\ldots,X_n$ are i.i.d. $B(1;p)$,

then $X=X_1+\ldots+X_n$ follows $B(n,p)$.

$E[X_i] = 1\,p + 0(1-p) = p$

$E[X] = E[\Sigma_i X_i] = \Sigma_i E[X_i] = \Sigma_i p = np$

Today's topic 2

Moment & Variance

Motivation

Consider the following three distributions.

Distr. 1.

- $\Pr[X=0] = 1/3$
- $\Pr[X=1] = 1/3$
- $\Pr[X=2] = 1/3$

Distr. 2.

- $\Pr[X=k] = 2^{-(k+1)}$

  for $k=0,1,2,\ldots$

Distr. 3.

- $\Pr[X=0] = 2/3$
- $\Pr[X=1] = 0$
- $\Pr[X=2^k] = 2^{-2k}$

  for $k=1,2,\ldots$

$\mathrm{Ex}[X] = 1$

$\mathrm{Ex}[X] = 1$

$\mathrm{Ex}[X] = 1$

Motivation

Consider the following three distributions.

Distr. 1.

- $\Pr[X=0] = 1/3$
- $\Pr[X=1] = 1/3$
- $\Pr[X=2] = 1/3$

$\mathrm{Ex}[X] = 1$

$\Pr[X>1] = 1/3$

$\Pr[X>2] = 0$

$\Pr[X>1000] = 0$

Distr. 2.

- $\Pr[X=k] = 2^{-(k+1)}$ for $k=0,1,2,\ldots$

$\mathrm{Ex}[X] = 1$

$\Pr[X>1] = 1/4$

$\Pr[X>2] = 1/8$

$\Pr[X>1000] = 1/512$

Distr. 3.

- $\Pr[X=0] = 2/3$
- $\Pr[X=1] = 0$
- $\Pr[X=2^k] = 2^{-2k}$ for $k=1,2,\ldots$

$\mathrm{Ex}[X] = 1$

$\Pr[X>1] = 1/3$

$\Pr[X>2] = 1/12$

$\Pr[X>1000] = 1/192$

Definitions

- ✓ k-th moment (k次の積率) of X

  $E[X^k]$

- ✓ variance (分散) of X

  $Var[X] := E[(X-E[X])^2]$

- ✓ standard deviation (標準偏差) of X

  $\sigma[X] := \sqrt{Var[X]}$

- ✓ covariance (共分散) of X, Y

  $Cov[X,Y] := E[(X-E[X])(X-E[Y])]$

Properties of Var and Cov

Thm.

$$\mathrm{Var}[X] = E[X^2] - (E[X])^2$$

$$\mathrm{Cov}[X,Y] = E[XY] - E[X]E[Y]$$

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y]$$

Properties of Var and Cov

Thm.

$$Var[X] = E[X^2] - (E[X])^2$$

$$Cov[X,Y] = E[XY] - E[X]E[Y]$$

$$Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y]$$

$$E[(X-E[X])^2] = E[X^2 - 2XE[X] + (E[X])^2]$$
$$= E[X^2] - 2E[X]E[X] + (E[X])^2$$
$$= E[X^2] - (E[X])^2$$

$$Cov[X, Y] = E[(X-E[X])(Y-E[Y])]$$
$$= E[(XY - XE[Y] - YE[X] + E[X]E[Y]]$$
$$= E[XY] - 2E[X]E[Y] + E[X]E[Y]$$
$$= E[XY] - E[X]E[Y]$$

Properties of Var and Cov

Thm.

$$\mathrm{Var}[X] = E[X^2] - (E[X])^2$$

$$\mathrm{Cov}[X,Y] = E[XY] - E[X]E[Y]$$

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y]$$

$$\begin{aligned}
\mathrm{Var}[X+Y] &= E[(X+Y)^2] - (E[X+Y])^2 \\
&= E[X^2+2XY+Y^2] - (E[X]+E[Y])^2 \\
&= E[X^2]-(E[X])^2 + E[Y^2]-(E[Y])^2 + 2E[XY]-2E[X]E[Y] \\
&= \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}[X,Y]
\end{aligned}$$

Properties of Var and Cov

Thm. Suppose X and Y are independent

$E[XY] = E[X]E[Y]$

$Cov[X,Y] = 0$

$Var[X+Y] = Var[X] + Var[Y]$

$$
\begin{aligned}
E[XY] &= \Sigma_x \Sigma_y\, xy \Pr[X=x \wedge Y=y] \\
&= \Sigma_x \Sigma_y\, xy \Pr[X=x] \Pr[Y=y] \\
&= (\Sigma_x\, x \Pr[X=x])(\Sigma_y\, y \Pr[Y=y]) \\
&= E[X]E[Y]
\end{aligned}
$$

$$
\begin{aligned}
Cov[X,Y] &= E[XY] - E[X]E[Y] \\
&= 0
\end{aligned}
$$

Properties of Var and Cov

Thm. Suppose X_i (i=1,…,n) are mutually independent

$$Var[X_1+X_2+\ldots+X_n] = Var[X_1] + Var[X_2] +\ldots +Var[X_n]$$

Linearity of independent variance: binomial distr.

Thm.

The variance of $X \sim B(n;p)$ is $np(1-p)$

proof

Suppose $X_1,\ldots,X_n$ are independent and identically distr. $B(1;p)$,

then $X=X_1+\ldots+X_n$ follows $B(n,p)$.

$E[(X_i)^2] = 1\ p + 0(1-p) = p$

$Var[X_i] = E[(X_i)^2] - (E[X_i]\ )^2 = p - p^2 = p(1-p)$

$Var[X] = Var[\Sigma_i\ X_i] = \Sigma_i\ Var[X_i] = \Sigma_i\ p = np(1-p)$