Lesson 4

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

Expectation (contd.)

Ex. Coupon collector

•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 ?

Ex. Coupon collector

•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 ?

Suppose you have already drawn $k-1$ kinds of coupon.

Let X_k denote the number of draws from $k-1$ to k.

- ➢ The probability is $p_k := \frac{n-(k-1)}{n}$
- ➢ The expected number is

$$\mathrm{E}[X_k] = \frac{1}{p_k} = \frac{n}{n-k+1}$$

Ex. Coupon collector

•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 ?

$$\mathrm{E}[X] = \mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \sum_{i'=1}^{n} \frac{1}{i'}$$

harmonic number

$$\ln n = \int_1^n \frac{1}{x}\mathrm{d}x \le \sum_{k=1}^{n} \frac{1}{k}$$

$$1 + \sum_{k=2}^{n} \frac{1}{k} \le 1 + \int_1^n \frac{1}{x}\mathrm{d}x = 1 + \ln n$$

Thm.

$$n \ln n \le E[X] \le n(1 + \ln(n))$$

Ex. Coupon collector

•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 ?
- ✓ What is the probability of completion after m trials?

Today's topic 1

Markov's inequality

Markov's inequality

Thm. Markov's inequality

Let X be a nonnegative random variable, then

$$\Pr[X \geq a] \leq \frac{\mathrm{E}[X]}{a}$$

holds for any a $>$ 0.

Markov's inequality

Thm. Markov's inequality

Let X be a nonnegative random variable, then

$$\Pr[X \geq a] \leq \frac{\mathrm{E}[X]}{a}$$

holds for any a > 0.

Proof.

$$\mathrm{E}\left[\frac{X}{a}\right] = \int_0^\infty \frac{x}{a} f(x)\mathrm{d}x = \int_0^a \frac{x}{a} f(x)\mathrm{d}x + \int_a^\infty \frac{x}{a} f(x)\mathrm{d}x$$
$$\geq \int_a^\infty \frac{x}{a} f(x)\mathrm{d}x \geq \int_a^\infty f(x)\,\mathrm{d}x = \Pr[X \geq a]$$

Thus,

$$\Pr[X \geq a] \leq \mathrm{E}\left[\frac{X}{a}\right] = \frac{\mathrm{E}[X]}{a} \qquad \square$$

Ex. Coupon collector

•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 ?
- ✓ What is the probability of completion after m trials?

rem.

$$n \ln n \leq E[X] \leq n(1 + \ln n)$$

too loose?

Using Markov's inequality,

$$\Pr[X \geq m] \leq \frac{E[X]}{m} \leq \frac{n(1 + \ln n)}{m}$$

e.g., n=100, m=1000,

$$\Pr[completion] \geq 1 - \Pr[X \geq 1001] \simeq 0.44$$

e.g., n=100, m=10000,

$$\Pr[completion] \geq 1 - \Pr[X \geq 10001] \simeq 0.94$$

Today’s topic 3

Chebyshev’s inequality

Chebyshev's inequality

Thm. Chebyshev's inequality

For any a $>$ 0.

$$\Pr[|X - \mathrm{E}[X]| \geq a] \leq \frac{\mathrm{Var}[X]}{a^2}$$

proof.

Remark that

$$\Pr[|X - \mathrm{E}[X]| \geq a] = \Pr[(X - \mathrm{E}[X])^2 \geq a^2]$$

Using Markov's inequality,

$$\Pr[(X - \mathrm{E}[X])^2 \geq a^2] \leq \frac{\mathrm{E}[(X - \mathrm{E}[X])^2]}{a^2} = \frac{\mathrm{Var}[X]}{a^2}$$

Chebyshev's inequality

Cor. Chebyshev's inequality

For any t > 0.

$$\Pr\left[X \geq (1+t)\mathrm{E}[X]\right] \leq \frac{\mathrm{Var}[X]}{(t\mathrm{E}[X])^2}$$

proof.

$$\begin{aligned}\Pr\left[X \geq (1+t)\mathrm{E}[X]\right] &= \Pr[X - \mathrm{E}[X] \geq t\mathrm{E}[X]] \\ &\leq \Pr\left[|X - E[X]| \geq t\mathrm{E}[X]\right] \\ &\leq \frac{\mathrm{Var}[X]}{(t\mathrm{E}[X])^2}\end{aligned}$$

Ex. Coupon collector

- •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 ?
- ✓ What is the probability of completion after m trials?

rem.

$$n \ln n \leq E[X] \leq n(1 + \ln n)$$

Using Markov's inequality,

$$\Pr[X \geq m] \leq \frac{E[X]}{m} \leq \frac{n(1 + \ln n)}{m}$$

too loose?

e.g., n=100, m=1000,

$$\Pr[completion] = 1 - \Pr[X \geq 1001] \simeq 0.44$$

e.g., n=100, m=10000,

$$\Pr[completion] = 1 - \Pr[X \geq 10001] \simeq 0.94$$

Ex. Coupon collector

•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 ?
- ✓ What is the probability of completion after m trials?

rem.

$$n \ln n \leq E[X] \leq n(1 + \ln n)$$

Using Chevyshev's inequality,

$$\Pr[X \geq (1+t)E[X]] \leq \frac{\mathrm{Var}[X]}{(t\mathrm{E}[X])^2}$$

Ex. Coupon collector

•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 ?
- ✓ What is the probability of completion after m trials?

Ex. 2.

$$\mathrm{Var}[X] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = \sum_{i=1}^{n} \frac{1-p_i}{p_i^2}$$

$$\leq \sum_{i=1}^{n} \frac{1}{p_i^2} = \sum_{i=1}^{n} \left(\frac{n}{n-i+1}\right)^2 = n^2 \sum_{i=1}^{n} \frac{1}{i^2} \leq n^2 \frac{\pi^2}{6}$$

Ex. Coupon collector

•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 ?
- ✓ What is the probability of completion after m trials?

rem.

$$n \ln n \leq E[X] \leq n(1 + \ln n)$$

still loose?
⇒ Chernoff's bound

Using Chevyshev's inequality,

$$\Pr[X \geq (1+t)E[X]] \leq \frac{\mathrm{Var}[X]}{(t\mathrm{E}[X])^2} \leq \frac{\frac{n^2\pi^2}{6}}{t^2(n \ln n)^2} = \frac{\pi^2}{6t^2(\ln n)^2}$$

e.g., n=100, m=1000 ($t \simeq \frac{m}{n \ln n} - 1 \simeq 1.1$),

Pr[Completion] ≥ 1-Pr[X ≥ 1000] ≈ 0.95

Law of Large number

Law of large numbers (大数の法則)

Def.

A series $\{Y_n\}$ converges Y in probability (Yに確率収束する), if

$$\forall \varepsilon > 0, \lim_{n \to \infty} \Pr[|Y_n - Y| < \varepsilon] = 1$$

independent and identically distributed (独立同一分布)

Thm. (law of large numbers; 大数の法則)

Let r.v. $X_1, \dots, X_n$ are i.i.d., w/ expectation μ, and variance σ^2,

then $Y_n := \frac{X_1 + \cdots + X_n}{n}$ converges μ in probability;

i.e.,

$$\forall \varepsilon > 0, \lim_{n \to \infty} \Pr\left[\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| < \varepsilon\right] = 1$$

Thm. (low of large numbers; 大数の法則)

Let r.v. $X_1, \ldots, X_n$ are i.i.d., w/ expectation μ, and variance σ^2,

then $Y_n := \frac{X_1 + \cdots + X_n}{n}$ converges μ in probability;

i.e.,

$$\forall \varepsilon > 0, \lim_{n \to \infty} \Pr\left[\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| < \varepsilon\right] = 1$$

$$\mathrm{E}[Y] = \mathrm{E}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{\mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]}{n} = \mu$$

$$\mathrm{Var}[Y] = \mathrm{Var}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{\mathrm{Var}[X_1] + \cdots + \mathrm{Var}[X_n]}{n^2} = \frac{\sigma^2}{n}$$

Thm. (low of large numbers; 大数の法則)

Let r.v. $X_1,\ldots,X_n$ are i.i.d., w/ expectation μ, and variance σ^2,

then $(X_1+\ldots+X_n)/n$ converge μ in probability;

i.e.,

$$\forall \varepsilon > 0, \lim_{n\to\infty} \Pr\left[\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| < \varepsilon\right] = 1$$

Using Chebyshev's inequality,

$$\Pr\left[\left|\frac{X_1+\cdots+X_n}{n} - \mu\right| \geq 1-\varepsilon\right] \leq \frac{\frac{\sigma^2}{n}}{(1-\epsilon)^2} \xrightarrow{n\to\infty} 0 \qquad \square$$

Recall

Thm. Chebyshev's inequality

For any $a > 0$.

$$\Pr[|X - \mathrm{E}[X]| \geq a] \leq \frac{\mathrm{Var}[X]}{a^2}$$

$$\mathrm{E}[Y] = \mu$$

$$\mathrm{Var}[Y] = \frac{\sigma^2}{n}$$