Statistik II - Exercise session 2
29.10.2014 & 5.11.2014

Info

- Classroom: SPA1 203
- Time: Wednesdays, 16:15 - 17:45
- in English
- Assignments on webpage (lvb>staff>PB)

Contact: Petra Burdejova
petra.burdejova@hu-berlin.de
Office: SPA1 R400 (upon agreement)

Schedule:

Date	Week	Exercises
15.10.14	E1	4-11, 4-18, 4-19
22.10.14	E1	4-11, 4-18, 4-19
29.10.14	E2	5-6, 5-10, 5-11, 5-14 ,5-16
5.11.14	E2	5-6, 5-10, 5-11, 5-14 ,5-16
12.11.14	E3	5-21, 5-23
19.11.14	E3	5-21, 5-23
26.11.14	E4	TBA
03.12.14	E4	TBA
10.12.14	E5	TBA
17.12.14	E5	TBA
07.01.15	E6	TBA
14.01.15	E6	TBA
21.01.15	E7	TBA
28.01.15	–	–
04.02.15	E8	Review for exam
11.02.15	E8	Review for exam

Review

- week 3 & week 4
- Slides: Distributions (04_Verteilungsmodelle) (1-50)

Distribution of discrete random variable

R.v. X having values $x_1, \dots, x_K$ with frequency $f(x_i)$ resp. probability p_i.

Probability mass function for discrete r.v.

$$\begin{aligned} \mathrm{P}(X = x_i) &= f(x_i) = p_i \qquad (i = 1, 2, \dots) \\ \mathrm{P}(a \le X \le b) &= \sum_{a \le x_i \le b} \mathrm{P}(X = x_i) = \sum_{a \le x_i \le b} f(x_i) \end{aligned}$$

Cumulative distr. function for discrete r.v.

$$F(x) = \mathrm{P}(X \le x) = \sum_{x_i \le x} f(x_i)$$

is monotoning increasing step function with jumps only at points $x_1, \dots, x_K$.

Parameters of discrete random variable

Expected value $\mathrm{E}(X) = \mu_X = \sum_{i=1}^{K} x_i \cdot f(x_i)$

Variance $\mathrm{Var}(X) = \sigma_X^2 = \sum_{i=1}^{K} (x_i - \mu_X)^2 \cdot f(x_i) = \sum_{i=1}^{K} x_i^2 \cdot f(x_i) - \mu_X^2$

Examples of Discrete distributions

- **Uniform:** ideal dice
- **Bernoulli:** 1 toss of coin
- **Binomial:** tosses of coin, lottery of colorful balls
- **Poisson:** occurrence of diseases, calls in center, rows in shop
- **Hypergeometric:** students chances for exam questions, quality of goods
- **Logarithmic:** claim frequency in insurance, purchased items by consumers

Notes: Poiss. dist. applied by L.v.Bortkiewicz investigating number of soldiers in army killed by horse kicks. Impact on reliability engineering..

Pascal's Triangle

1. Set 1 on edges
2. Sum number above to get number below
3. Bin. coefficients of 5th order on 5th row

Overview of discrete distributions

Distribution $\mathcal{L}(X)$	PMF $\mathsf{P}(X = x_i)$	Mean $\mathrm{E}\left[X\right]$	Variance $\mathsf{Var}\left[X\right]$	PGF $\mathsf{P}_X(s)$	CDF F(x)
Uniform	$\frac{1}{K}$ $i \in \{0, \ldots, K\}$	$\frac{\sum_{i=1}^K x_i}{K} = \mu_x$	$\frac{1}{K} \sum_{i=1}^K (x_i - \mu_x)^2$	$\frac{1-s^{K+1}}{K(1-s)}$	$F(x) = \begin{cases} 0 & x < x_1 \\ \frac{i}{K} & x_i \leq x < x_{i+1} \\ 1 & x \geq x_K \end{cases}$
Bernoulli	$p^x(1-p)^{1-x}$ $x \in \{0, 1\}$	p	$p(1-p)$	$1 - p + ps$	$F(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$ $x \in \{0, 1, \ldots, n\}$	np	$np(1-p)$	$(1 - p + ps)^n$	$\sum_{k=0}^x \binom{n}{k} p^k (1-p)^{n-k}$
Poisson	$\frac{\lambda^x}{x!} e^{-\lambda}$ $x \in \{0, 1, \ldots\}$	λ	λ	$e^{\lambda(s-1)}$	$\sum_{k=0}^x \frac{\lambda^k}{k!} e^{-\lambda}$
Hyper-geometric	$\frac{\binom{M}{x} \cdot \binom{N-M}{N-x}}{\binom{N}{n}}$ $x \in \{0, 1, \ldots \min(n, M)\}$	$n \cdot \frac{M}{N}$	$n \cdot \frac{M}{N} \left(1 - \frac{M}{N}\right) \left(\frac{N-n}{N-1}\right)$	(*)	(*)
Logarithmic	$-\frac{p^x}{x \log(1-p)}$ $x \in \{1, 2, \ldots\}$	$\frac{p}{(p-1)\log(1-p)}$	$-\frac{p(p+\log(1-p))}{(1-p)^2[\log(1-p)]^2}$	$\frac{\log(1-ps)}{\log(1-p)}$	(*)

(*) formulas can be found on Wikipedia or Wolfram Alpha , ...

Proof that the hypergeometric distribution with large N approaches the binomial distribution.

4

I have this problem on a textbook that doesn't have a solution. It is:

Let

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} ,$$

and keep $p = \frac{r}{N}$ fixed. Prove that

$$\lim_{N\to\infty} f(x) = \binom{n}{x} p^x (1-p)^{n-x} .$$

Although I can find lots of examples using the binomial to approximate the hypergeometric for very large values of N, I couldn't find a full proof of this online.

(probability) (statistics)

share improve this question

edited Mar 14 '13 at 19:28
Brian M. Scott
244k ● 24 ■ 249 ▲ 513

asked Mar 14 '13 at 19:17
user54609
810 ■ 3 ▲ 22

add a comment

1 Answer

active oldest votes

2

Write the pmf of the hypergeometric distribution in terms of factorials:

$$\begin{aligned}\frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} &= \frac{r!}{x!\cdot(r-x)!}\frac{(N-r)!}{(n-x)!\cdot(N-n-(r-x))!}\cdot\frac{n!\cdot(N-n)!}{N!} \\ &= \binom{n}{x}\cdot\frac{r!/(r-x)!}{N!/(N-x)!}\cdot\frac{(N-r)!\cdot(N-n)!}{(N-x)!\cdot(N-r-(n-x))!} \\ &= \binom{n}{x}\cdot\frac{r!/(r-x)!}{N!/(N-x)!}\cdot\frac{(N-r)!/(N-r-(n-x))!}{(N-n+(n-x))!/(N-n)!} \\ &= \binom{n}{x}\cdot\prod_{k=1}^{x}\frac{(r-x+k)}{(N-x+k)}\cdot\prod_{m=1}^{n-x}\frac{(N-r-(n-x)+m)}{(N-n+m)}\end{aligned}$$

Now taking the large N limit for fixed r/N, n and x we get the binomial pmf, since

$$\lim_{N\to\infty}\frac{(r-x+k)}{(N-x+k)} = \lim_{N\to\infty}\frac{r}{N} = p$$

and

$$\lim_{N\to\infty}\frac{(N-r-(n-x)+m)}{(N-n+m)} = \lim_{N\to\infty}\frac{N-r}{N} = 1-p$$

share improve this answer

answered Mar 14 '13 at 19:34
Sasha
49.2k ● 4 ■ 63 ▲ 135

asked 1 year ago
viewed 3279 time
active 11 months

Linked

0 Choosing right type of variable: binomial or p (Probability question)

-1 Help following a proof regarding hypergeome distribution

Related

1 Why Binomial Distribut formula includes the "r happening" probability

2 approximation hypergeometric distrib with binomial

0 what's the distribution inverse of a random va that follows a negative binomial distribution?

2 Bernoulli distribution: expectation problem w independent random v

3 Approximating hypergeometric distrib with poisson

2 Confidence Interval for Binomial

2 Poisson Distribution - Questions about λ par

0 Using binomial distribu and continuity correcti Please help.

0 Binomial distribution concept?

Source: `http://math.stackexchange.com/questions/330553`

Distribution of continuous random variable

Probability mass function for continuous r.v.

$$P(a < X \le b) \quad = \quad \int_a^b f(x)\,dx \text{ for all } a, b \text{ such that } a \le b$$

Cumulative distr. function for continuous r.v.

$$F(x) \quad = \quad P(-\infty < X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

Parameters of continuous random variable

Expected value $\mathrm{E}(X) = \mu_X = \int_{-\infty}^{+\infty} x \cdot f(x)dx$

Variance $\mathrm{Var}(X) = \sigma_X^2 = \int_{-\infty}^{+\infty} (x - \mu_X)^2 \cdot f(x)\,dx = \int_{-\infty}^{+\infty} x^2 \cdot f(x)\,dx - \mu_X^2$

Overview of continuous distributions

Distribution $\mathcal{L}(X)$	PDF $f(x)$	CDF F(x)	Mean $\mathrm{E}[X]$	Variance $\mathsf{Var}[X]$
Uniform	$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & otherwise \end{cases}$	$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x < b \\ 1 & b \le x \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential	$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} & x \ge 0, \lambda > 0 \\ 0 & x < 0 \end{cases}$	$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \ge 0 \\ 0 & x < 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal	$\frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\cdot\sigma^2}\right)$	$\int_{-\infty}^{x} f(t)\,dt$	μ	σ^2

Examples of Continuous distributions

- **Uniform:** waiting time
- **Exponential:** interval between Poisson events (bulbs, machines defects,..)
- **Normal:** observational errors

Normal distribution:

- There is standard normal distribution $N(0;1)$ with cdf $\Phi(z)$
- If $X \sim N(\mu;\sigma)$ then $\frac{x-\mu}{\sigma} \sim N(0;1)$
- If X and Y are jointly normally distributed, $X \sim N(\mu_X;\sigma_X)$ and $Y \sim N(\mu_Y;\sigma_Y)$ then $X + Y \sim N(\mu_X + \mu_Y; \sigma_{X,Y})$ where $\sigma_{X,Y} = \sqrt{\sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y}$

Exercises

Exercise 5-6 - Formal Errors

In a company 10% of the existing (large numbers of) documents have formal error. Auditor will select 10 documents randomly.

a) What is the distribution of the number of faulty documents?

b) Calculate the probability that he finds more than one incorrect document.

Exercise 5-10 - Eggs

It is known that there are 2 faulty eggs in every 6-pack. We will take 3 eggs from pack and try them (i.e. throw them on pan).

a) What is the probability, that exactly one is faulty?

b) What is the probability, that exactly one is faulty at most?

c) What is the probability, that exactly three eggs are faulty?

d) How many faulty eggs can man expect from these 3 chosen ones?

At one small farm are produced 500 eggs in long time-period. It is know that with 80% probability such egg is not faulty. Order of 20 eggs was made (random choice of eggs).

e) What is the probability approximately, that more than 2 eggs in order are faulty?

f) Compute the expected value of ”number of good eggs in order”.

g) Compute approximated probability, that there are exactly 16 faulty eggs in order.

Exercise 5-11 - Phone calls

There are on average 2,5 number of calls per minute in firm between 2pm and 4pm.

a) What is the distribution of r.v. ”Number of received calls per minute”?

b) What is the probability, that during exact minute (in this period)

- no
- less than three
- four or more

calls are received?

Exercise 5-14 - Electronic component

For one electronic component we can expect 48 fallouts per day (24hrs). Fallouts are short, random and independent on each other.

a) What is the distribution of time between two fallouts? State the type and parameter.

b) What is the probability that it will take more than 2 hours till next fallout occurs?

c) Formulate (explain) following form using given example.

$$\int_1^2 2e^{-2x}dx \tag{1}$$

d) Assume that electronic system consists of two such components, which are functional independently on each other. System fails when when one of components does not work. What is the probability, that system works more than 2 hours?

Exercise 5-16 - Steel pins

Machine produces steel pins. Unfortunately, diameter of pins fluctuates. R.v. "diameter of a pin" X_1 is normally distributed with $\mu = 6$mm and $\sigma = 0.4$mm.

a) What is the probability that diameter of pin is more than 2% away from $\mu = 6$mm?

b) What is the probability that diameter of pin is exactly 6mm?

c) Which value will not be exceeded with probability of 85%?

Another machine, which works independently on the first one, drilling holes into a work piece, in which the steel pins will be used. Also the diameter of holes is normally distributed random variable X_2 with $\mu = 6.05$mm and $\sigma = 0.3$mm.

d) What is the probability that diameter of hole is less than 6mm?

e) What is the distribution of $Y = X_1 - X_2$?

f) What is the probability that pin does not fit in the hole?