FEGYVERNEKI SÁNDOR,

PROBABILITY THEORY AND MATHEMATICAL STATISTICS

5

Új Magyarország FEJLESZTÉSI TERV

A Műszaki Földtudományi Alapszak tananyagainak kifejlesztése a TÁMOP 4.1.2-08/1/A-2009-0033 pályázat keretében valósult meg.

V. Inequalities and Basic Limit Theorems

1. Inequalities and Basic Limit

Theorem: (Markov's Inequality) Let X be a random variable with expectation $E(X)$ and such that $P(X<0)=0$. Then, for every $t>0$,

$$P(X\geq t)\leq\frac{E(X)}{t}.$$

Theorem: (Chebyshev's Inequality) Let X be a random variable with expectation $E(X)$ and standard deviation σ. Then, for each $t>0$,

$$P(|X-E(X)|\geq t)\leq\frac{\sigma^2}{t^2}.$$

Theorem: (One-Sided Inequality) Let X be a random variable with expectation $E(X)$ and variance σ^2. Then,

$$P(X\leq t)\leq\frac{\sigma^2}{\sigma^2+(t-E(X))^2}\quad\text{if}\quad t<E(X),$$

or

$$P(X>t)\leq\frac{\sigma^2}{\sigma^2+(t-E(X))^2}\quad\text{if}\quad t>E(X).$$

Theorem: (Weak Law of Large Numbers) Let $X_1, X_2, \ldots$ be independent random variables on (Ω, Σ, P) with common distribution (i.e. $P(X_n<x)=F(x),\quad x\in\mathbf{R},$ for $n\geq 1$) and with one moment finite. If $S_n=\sum_{k=1}^{n}X_k$, then

$$\lim_{n\to+\infty}P\left(\left|\frac{S_n-E(S_n)}{n}\right|\geq\varepsilon\right)=0.$$

Corollary: Let A be an event and S_n the number of times that A occurs in a Bernoulli sequence of n trials. Then, for each $\varepsilon > 0$,

$$\lim_{n\to+\infty} P(|\frac{S_n}{n} - P(A)| \geq \varepsilon) = 0.$$

Theorem: (Strong Law of Large Numbers) Let $X_1, X_2, \ldots$ be independent random variables on (Ω, Σ, P) with common distribution and $S_n = \sum_{k=1}^{n} X_k$. Then

$$P(\{\omega : \lim_{n\to+\infty} \frac{S_n(\omega)}{n} = \alpha\}) = 1 \quad iff \quad E(|X_1|) < \infty.$$

Theorem: (Central Limit Theorem) Let $X_1, X_2, \ldots$ be independent, identically distributed random variables each with expectation μ and variance $\sigma^2 > 0$. If $S_n = \sum_{k=1}^{n} X_k$, then

$$\lim_{n\to+\infty} P(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} < x) = \Phi(x), \qquad x \in \mathbf{R},$$

where Φ is the standard normal distribution function.

Theorem: (Moivre-Laplace Theorem) The probability that a random variable with distribution $b(n,p)$ will have an observed value lying between a and b, inclusive, for any two integers a and b, is given approximately by

$$\sum_{k=a}^{b} \binom{n}{k} p^k q^{n-k} = \Phi(\frac{b - np + \frac{1}{2}}{\sqrt{npq}}) - \Phi(\frac{a - np - \frac{1}{2}}{\sqrt{npq}}).$$

PROBLEM

EXAMPLE 1 How many trials of an experiment with two outcomes, called A and B, should be performed in order that the probability be 0.95 or better that the observed relative frequency of occurrences of A will differ from the probability p of occurrence of A by no more than 0.02?

Solution:

(i) Apply the Chebyshev's inequality

$$P(|X - E(X)| \geq t) \leq \frac{\sigma^2}{t^2}.$$

Denote f_n the relative frequency, then

$$P(|f_n - p| < t) \geq 1 - \frac{p(1-p)}{nt^2}$$

since $Var(f_n) = \frac{p(1-p)}{n}$. But $p(1-p) \leq \frac{1}{4}$, thus

$$1 - \frac{1}{4nt^2} \geq 0.95.$$

Because $t = 0.02$, therefore

$$n \geq \frac{1}{4 \cdot 0.02^2(1-0.95)} = 12500.$$

(ii) Apply the central limit theorem. If p is the probability of success at each trial then the number of successes S_n in n independent repeated Bernoulli trials approximately satisfies, for any $x > 0$,

$$P\left(\frac{|S_n - np|}{\sqrt{npq}} < x\right) = 2\Phi(x) - 1.$$

Let $x = t\sqrt{\frac{n}{pq}}$. Consequently,

$$P(|f_n - p| < t) = 2\Phi\left(t\sqrt{\frac{n}{pq}}\right) - 1.$$

Define $K(\alpha)$ as the solution of the equation

$$2\Phi(K(\alpha)) - 1 = \alpha.$$

Thus we obtain the conclusion that

$$P(|f_n - p| < t) \geq \alpha, \quad if \quad t\sqrt{\frac{n}{pq}} \geq K(\alpha).$$

Since $pq \leq \frac{1}{4}$ for all $0 \leq p \leq 1$

$$n \geq \frac{K^2(\alpha)}{4t^2}.$$

If $\alpha = 0.95$ and $t = 0.02$, then n should be chosen so that $n \geq 2500$.

Thus the number of trials required for f_n to be within 0.02 of p with probability greater than 0.95 is approximately 2500, which is $\frac{1}{5}$ of the number of trials that Chebyshev's inequality states requires.

PROBLEM

EXAMPLE 2 In 40000 independent tosses of a coin heads appeared $20,400$ times. Find the probability that if the coin were fair one would observe in $40,000$ independent tosses (i) $20,400$ or more heads, (ii) between $19,600$ and $20,400$ heads.

Solution:

Let X be the number of heads in $40,000$ independent tosses of a fair coin. Then X obeys a binomial probability law with mean $np = 20,000$, variance $npq = 10,000$, and standard deviation 100. Apply the Moivre-Laplace theorem:

$$P(X > 20,400) = 0.0001 \quad \text{and} \quad P(19,600 < X < 20,400) = 0.99995$$

approximately.

2. Exercises

Solutions: ⊙ visible ○ invisible

1. Let X be a normally distributed random variable with parameters (i) $m = 0$ and $\sigma = 1$, (ii) $m = 0$ and $\sigma = 2$. For α in $0 < \alpha < 1$ define $J(\alpha)$ and $K(\alpha)$ so that

$$P(X > J(\alpha)) = \alpha, \qquad P(|X| < K(\alpha)) = \alpha.$$

Find $J(\alpha)$ and $K(\alpha)$ for $\alpha = 0.05,\ 0.10,\ 0.50,\ 0.90,\ 0.95,\ 0.99.$

Solutions:

(i)

α	0.05	0.10	0.50	0.90	0.95	0.99
$J(\alpha)$	1.645	1.282	0.000	-1.282	-1.645	-2.326
$K(\alpha)$	0.063	0.126	0.675	1.645	1.960	2.576

(ii)

α	0.05	0.10	0.50	0.90	0.95	0.99
$J(\alpha)$	3.290	2.564	0.000	-2.564	-3.290	-4.652
$K(\alpha)$	0.126	0.252	1.350	3.290	3.920	5.152

2. Assume that the height in centimeters of a man aged 21 is a normally distributed random variable with parameters $m = 170$ and $\sigma = 5$. What is the probability that the height of a man aged 21 will be greater than 170 centimeters, given that it is greater than 160 centimeters?

Solution:
0.512.

3. A machine produces bolts in a length (in inches) found to obey a normal probability law with parameters $m = 10$ and $\sigma = 0.10$. The specifications for the bolt call for items with a length (in inches) equal to 10.05 ± 0.12. A bolt not meeting these specifications is called defective.

i. What is the probability that a bolt produced by this machine will be defective?

ii. If the machine were adjusted so that the length of bolts produced by it is normally distributed with parameters $m = 10.10$ and $\sigma = 0.10$, what is the probability that a bolt produced by the machine will be defective?

iii. If the machine is adjusted so that the lengths of bolts produced by it are normally distributed with parameters $m = 10.05$ and $\sigma = 0.06$, what is the probability a bolt produced by the machine will be defective?

Solutions:
(i) 0.2866, (ii) 0.2866, (iii) 0.0426.

---

4. Suppose X and Y are independent random variables such that $E(X) = E(Y) = 1$ and $Var(X) = Var(Y) = 1$.

i. Using Chebyshev's inequality, give a bound on the probability that $|X - Y| > 6$.

ii. Using Chebyshev's inequality, find a bound on the probability that $|X - 1| > \frac{1}{2}$.

iii. Compare the bound found in (ii) with exact values in the cases when X is known to be exponentially and normally distributed.

Solutions:
(i) $\frac{1}{18}$, (ii) 1, (iii) 0.6166, 0.617.

---

5. It is known that bacteria of a certain kind occur in water at the rate of two bacteria per cubic centimeter of water. Assuming that this phenomenon obeys a Poisson probability law, what is the probability that a sample of two cubic centimeters of water will contain

i. no bacteria,

ii. at least two bacteria?

Solutions:
$\lambda = 4$, (i) e^{-4}, (ii) $1 - 5e^{-4}$.

---

6. In a certain published book of 520 pages 390 typographical errors occur. What is the probability that four pages, selected randomly by the printer as examples of his work, will be free from errors?

Solution:
$\lambda = 3$, e^{-3}.

---

7. Suppose an on-line computer system is proposed for which it is estimated that the mean response time $E(T)$ is 10 seconds.

i. Use Markov's inequality to estimate the probability that the response time T will be 20 seconds or more.

ii. It is estimated that the standard deviation of response time is 2 seconds. Use Chebyshev's inequality to estimate the probability that the response time will be between 4 and 16 seconds.

Solutions:
(i) $P(T \le 20) \le \frac{1}{2}$, (ii) $P(4 < T < 16) = \frac{8}{9}$.

---

8. A mathematical model of a proposed on-line computer system gives a mean time to retrieve a record from a direct access storage device of 400 milliseconds with standard deviation of 116 milliseconds. One design criterion requires that 90% of all retrieval times must not exceed 750 milliseconds. Use the one-sided inequality to test the design criterion.

> **Solution:**
> $P(T < 750) \leq 0.9013.$

---

9. A certain access method, called method A, has been found to give a mean record retrieval time of 36 milliseconds with a standard deviation of 7 milliseconds, while method B has a mean retrieval time of 42 milliseconds with a standard deviation of 4 milliseconds. If a major design objective is to have 90% of all individual retrievals completed in 55 milliseconds or less, which method should be selected?

> **Solution:** Method B.

---

10. Let X be a Poisson distributed random variable with parameter $c.$ Prove that

i. $P(X \leq \frac{c}{2}) \leq \frac{4}{c+4} < \frac{4}{c},$

ii. $P(X \geq 2c) \leq \frac{1}{1+c} < \frac{1}{c}.$

> ***Hint:*** Use one-sided inequality.

---

11. Use Chebyshev's inequality to determine how many times a fair coin must be tossed in order that the probability will be at least 0.90 that the ratio of the observed number of heads to the number of tosses will lie between 0.4 and $0.6.$

> **Solution:** 250.

---

12. Consider a group of N men playing the game of "odd man out" (that is, they repeatedly perform the experiment in which each man independently tosses a fair coin until there is an "odd" man, in the sense that either exactly 1 of the N coins falls heads or exactly 1 of the N coins falls tails). Find, for (i) $N = 4,$ (ii) $N = 8,$ the exact probability that the number of repetitions required to conclude the game will be within 2 standard deviations of the mean number of repetitions required to conclude the game. Compare your answer with the lower bound given by Chebyshev's inequality.

> **Solutions:**
> (i) $1 - \frac{1}{2^4},$ (ii) $1 - \left(\frac{15}{16}\right)^{47},$ Chebychev bound 0.75.

---

13. A sample is taken to find the proportion p of smokers in a certain population. Find a sample size so that the probability is (i) 0.95 or better, (ii) 0.99 or better that the observed proportion of smokers will differ from the true proprotion of smokers by less than (a) $1\%,$ (b) $10\%.$

> **Solution:**
> Chebyshev bound, (i) (a) 50 000, (b) 500, (ii) (a) 250 000, (b) 2500.

Normal approximation, (i) (a) 9600, (b) 96, (ii) (a) 16 000, (b) 166.

14. If you wish to estimate the proportion of engineers and scientists who have studied probability theory and you wish your estimate to be correct, within 2%, with probability 0.95 or better, how large a sample should you take

i. if you feel confident that the true proportion is less than 0.2,

ii. if you have no idea what the true proportion is.

Solutions:
Chebyshev: (i) 8000, (ii) 12000. Normal: (i) 1537, (ii) 2400.

15. In $10,000$ independent tosses of a coin $5,075$ heads were observed. Find approximately the probability of observing (i) exactly $5,075$ heads, (ii) $5,075$ or more heads if the coin (a) is fair, (b) has probability 0.51 of falling heads up.

Solutions:
(i) (a) 0.003, (b) 0.008, (ii) (a) 0.068, (b) 0.695.

16. Find the probability that in $3,600$ independent repeated trials of an experiment, in which the probability of success of each trial is p, the number of success is between $3,600p-20$ and $3,600p+20$, inclusive, if (i) $p=\frac{1}{2}$, (ii) $p=\frac{1}{3}$.

Solutions:
(i) 0.506, (ii) 0.532.

17. Suppose that (i) 2, (ii) 3 restaurants compete for the same 800 patrons. Find the number of seats that each restaurant should have order to have a probability greater than 95% that it can serve all patrons who come to it (assuming that all patrons arrive at the same time and choose, independently of one another, each restaurant with equal probability).

Solutions:
423, (ii) 289.

18. Suppose that among 10000 students at a certain college 100 are red-haired.

i. What is the probability that a sample of 100 students, selected with replacement, will contain at least one red-haired student?

ii. How large is a random sample, drawn with replacement, if the probability of it containing a red-haired student is 0.95?

It would be more realistic to assume that the sample is drawn without replacement. Would the answers to (i) and (ii) change if this assumption were made?

Hint: State conditions under which the hypergeometric distribution is approximated by Poisson distribution.

Solution:
(i) $1-\frac{1}{e}$, (ii) 300.

19. Assuming that intelligence quotients are normally distributed with mean 100 and standard deviation 15, calculate:

i. the probability that a person chosen at random has an I.Q. between 88 and 118,

ii. the percentage of the population with I.Q. greater than 130.

Solutions:
(i) 0.673, (ii) 0.0228, i.e. 2.3%.

---

20. Which of the following sets of evidence throws more doubt on the hypothesis that new born babies are as likely to be boys and girls:

i. of 10000 new born babies, 5100 are male,

ii. of 1000 new born babies, 510 are male.

Solutions:
(i) throws more doubt than (ii).

---

21. As an estimate of the unknown mean $E(X)$ of a random variable, it is customary to take the sample mean $\overline{X} = (X_1 + ... + X_n)/n$ of a random sample $X_1, ..., X_n$ of the random variable X. How large a sample should one observe if there is to be a probability of at least 0.95 that the sample mean $\overline{X}$ will not differ from the true mean $E(X)$ by more than 25% of the standard deviation?

Solution: 62.

---

22. Consider a game of chance in which one may win 10 dollars or lose $1, 2, 3,$ or 4 dollars; each possibility has probability 0.20. How many times can this game be played if there is to be a probability of at least 95% that in the final outcome the average gain or loss per game will be between -2 and $+2$?

Solution: 25 or more.

---

23. Add 100 real numbers, each of which is rounded off to the nearest integer. Assume that each rounding-off error is random variable uniformly distributed between $-\frac{1}{2}$ and $\frac{1}{2}$ and that the 100 rounding-off errors are independent. Find approximately the probability that the error in the sum will be between -3 and 3. Find the quantity A that the probability is approximately 99% that the error in the sum will be less than A in absolute value.

Solution: 0.70, 7.4.

---

24. A delivery truck carries loaded cartons of items. If the weight of each carton is a random variable, with mean 50 pounds and standard deviation 5 pounds, how many cartons can the truck carry so that the probability of the total load exceeding 1 ton will be less than 5% ?

Solution: 38.

---

25. A random variable X has an unknown mean and known variance 4. How large a random sample should one take if the probability is to be at least 0.95 that the sample mean will not differ from the true mean $E(X)$ by

i. more than 0.1,

ii. more than 10% of standard deviation of X,

iii. than 10% of the true mean of X, if the true mean of X is known to be greater than 10?

Solutions:
(i) $n \geq 1537$, (ii) $n \geq 385$, (iii) $n \geq 16$.,

---

26. Bank tellers in certain bank make errors in entering figures in their ledgers at the rate of 0.75 errors per page of entries. What is the probability that in 4 pages there will be 2 or more errors?

Solution: 0.8008.

---

27. A radioactive source is observed during 4 time intervals of 6 seconds each. The number of particles emitted during each period are counted. If the particles emitted obey a Poisson probability law, at a rate of 0.5 particles emitted per second, find the probability that

i. in each of the 4 time intervals 3 or more particles will be emitted,

ii. in at least 1 of the 4 time intervals 3 or more particles will be emitted.

Solutions:
(i) 0.109, (ii) 0.968.

---

28. Suppose that the customers enter a certain shop at the rate of 30 persons an hour.

i. What is the probability that during a $2-$ minute interval either no one will enter the shop or at least 2 persons will enter the shop?

ii. If you observed the number persons entering the shop during each of 30 $2-$ minute intervals, would you find it suprising that 20 or more of these intervals had the property that either no one or at least 2 persons entered the shop during that time?

Solutions:
(i) 0.632, (ii) not suprising, since the number of 1-minute intervals in an hour in which either no enters or 2 or more enter obeys a binomial probability law with mean 18.9 and variance 6.96.

---

29. In a large fleet of delivery trucks the average number inoperative on any day because of repairs is 2. Two standby trucks are available. What is the probability that on any one day

i. no standby trucks will be needed,

ii. the number of standby trucks is inadequate?

Solutions:
(i) 0.1353, (ii) 0.3233.

---

30. Consider a restaurant located in the business section of a city. How many seats should it have available if it wishes to serve at least 95% of all those who desire its services in a given hour, assuming that potential customers (each of whom takes at least an hour to eat) arrive in accord with the following schemes:

i. 1000 persons pass by the restaurant in a given hour, each of whom has probability $\frac{1}{100}$ of desiring to eat in the restaurant (that is, each person passing by the restaurant enters the restaurant once in every 100 times),

ii. persons, each of whom has probability $\frac{1}{100}$ of desiring to eat in the restaurant, pass by the restaurant at the rate of 1000 an hour,

iii. persons desiring to be patrons of the restaurant, arrive at the restaurant at the rate of 10 an hour.

Solution: 15.

Digitális Egyetem, Copyright © Fegyverneki Sándor, 2011