情報学基礎 B

統計 2
Binomial, Poisson, Normal Distribution

Binomial Distribution(二項分布):
We are interested in the number of times an event A occurs in n independent trials. In each trial the event A has the same probability of occurrence $P(A) = p$. Then, in a trial, the probability that A will not occur is $q = (1-p)$. In n trials the random variable that interests us is $X =$ *Number of times A occurs in n trials.*
Thus, $X = x$ means that event of interest A occured in x trials and it did not occur in $n-x$ trials. They may look as follows.

$$\underbrace{AA\cdots A}_{x\ times} \qquad \underbrace{BB\cdots B}_{n-x\ times}$$
$$or \quad ABBAAAB\cdots B \qquad (A:\ x\ times\ \ B:\ n-x\ times)$$

This may happen in

$$\frac{n!}{x!(n-x)!} = \begin{pmatrix} n \\ x \end{pmatrix} \qquad number\ of\ ways$$

The probability is $P(X = x)$ of $X = x$, of obtaining precisely x A's in n trials. Hence X has the probability density function

$$f(x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x q^{n-x} \qquad (x = 0, 1, \cdots\ n)$$

The above is known as binomial distribution or bernoulli distribution(ベルヌーイ分布)
The occurrence of an event is called success (成功)(what it actually means is the event of interest. Thus, it may mean that you miss your train, or a product from an assembly line is faulty).
Nonoccurrence of an event is called failure(失敗).
Mean of a binomial distribution:

$$\mu = np$$

variance and standard deviation of a binomial distribution:

$$\sigma^2 = npq = np(1-p) \quad and\ therefore \quad \sigma = \sqrt{np(1-p)}$$

Example: Compute the probability of obtaining at least two "Six"es in rolling a fair die 4 times.
(4 回サイコロを投げて少なくとも 2 回以上 6 が出る確率を求めよ)

Poisson Distribution(ポアソン分布)

Poisson distribution is another discrete probability distribution. But here the success probability, p, is very low, and number of trials is high. In such a case, probability calculation using Binomial distribution is computationally heavy. Poisson distribution gives a more compact expression when $p \to 0$ and $n \to \infty$. The probability of x successes in n trials, where x could be $0, 1, 2, \cdots, n$ is given by

$$f(x) = \frac{\mu^x}{x!} e^{-\mu} \quad \textit{where } \mu = np$$

Poisson distribution follows from the fact that,

$$\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x$$

The mean and variance of Poisson distribution is same as in Binomial distribution, i.e.,

$$\mu = np \quad and \quad \sigma = \sqrt{np(1-p)} \approx \sqrt{np} \quad as \quad p \to 0$$

Examples of Poisson distribution:

1. If the probability of producing a defective screw is $p = 0.01$, what is the probability that a lot of 100 screws will contain more than 2 defectives ?
(欠陥のあるねじがある確率が $p = 0.01$ のとき、100 個のねじの中に 2 つ以上欠陥のあるねじが存在する確率を求めよ)

2. The mean number of misprints per page in a book is 1.2. What is the probabiity of finding on a particular page (a) no misprints; (b) three or more misprints? (平均的 1 ページにある印刷誤りは 1.2 です。ある 1 ページに（あ）誤りが一つもないと (い) 三つ以上の誤りがあるの確立を求めよ。

Problems:

1. During 8:00 am ∼ 6:00 pm total incoming call is 10. What is the probability of 1 call between 9:00 am to 10:00 noon? What is the probability of more than 2 calls between 9:00 am to 12:00 noon ?
(8:00 am ∼ 6:00 pm のあいだに 10 回電話がかかってくるときに、9:00 am から 10:00 pm の間に 1 回だけ電話がかかってくる確率、および 9:00 am から 12:00 noon までに 2 回電話がかかってくる確率をそれぞれ求めよ)

2. The following table shows us the death toll in a day during 1096 days. What is the probability of 2 people die on a single day? What is the probability that more than 2 people die on a single day?
(下の表は 1096 日間の死亡者数である。2 人が 1 日で死んでしまう確率、と 2 人以上が 1 日で死んでしまう確率をそれぞれ求めよ)

表 1:

Death toll is	0	1	2	3	4	5	6	7	8	9	10
Number of days	162	267	271	185	111	61	27	8	3	1	0

3. If on the average, 2 cars enter a certain parking lot per minute. what is the probability that during any given minute 4 or more cars will enter the lot? (駐車場に一分間平均 2 台の車が入ってくるとき、1 分間に駐車場に 4 台以上の車が入ってくる確率は?)

4. Five fair coins tossed simultaneously. Find the probability function of the random variable $X = Number\ of\ heads$ and compute the probabilities of obtaining no heads, precisely 1 head, at least 1 head, not more than 4 heads. (5 つのコインが同時に投げられるとき、確率変数 $X = Number\ of\ heads$ の確率関数を求め、全て裏、1 つだけ表、少なくとも一つは表、4 つ以上表、である確率を求めよ)

5. If the probability of hitting a target is 25% and 4 shots are fired independently, what is the probability that the target will be hit at least once? (ターゲットに当たる確率が 25%で、それぞれ独立に 4 発撃つとき、ターゲットに少なくとも 1 発以上当たる確率を求めよ)

6. In Prob.5, if the probability of hitting would be 5% and we fired 20 shots, would the probability of hitting at least once be less than, equal to, or greater than that in Prob.2? Guess first. (Prob.2 において、5% の命中率で 20 回撃つとき、ターゲットに少なくとも 1 発以上当たる確率を求めよ)

7. Classical experiments by E.Rutherford and H.Geiger in 1910 showed that the number of alpha particles emitted per second in a radioactive process is a random variable X having a Poisson distribution. If X has mean 0.5, what is the probability of observing two or more particles during any given second? (放射性過程において、1 秒単位で放たれたα粒子の数は、ポアソン分布をもつ X の確率変数に等しい。X の平均が 0.5 のとき、α粒子が 2 つ以上観測できる確率をもとめよ)

Normal Distribution(標準分布)

The normal distribution or Gauss distribution is defined as the distribution with the density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)\right]$$

The distribution function

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} exp\left[-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2\right]\ dv$$

Standardized normal distribution with mean 0 and standard deviation 1 we denote $F(x)$ by

$\Phi(z)$. Then we simply have from distribution function

$$\Phi(z) \;=\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\, du$$

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

The probability that a normal random variable X with mean μ and standard deviation σ assume any value in an interval $a \;<\; x \;\leq\; b$ is

$$P(a \;<\; x \;\leq\; b) = F(b) \;-\; F(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) \;-\; \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$\begin{aligned}
P(\mu-\sigma \;< X \;\leq\; \mu+\sigma) \;&\approx\; 68\% \\
P(\mu-2\sigma \;< X \;\leq\; \mu+2\sigma) \;&\approx\; 95.5\% \\
P(\mu-3\sigma \;< X \;\leq\; \mu+3\sigma) \;&\approx\; 97.7\%
\end{aligned}$$

$$\begin{aligned}
P(\mu-1.96\sigma \;< X \;\leq\; \mu+1.96\sigma) \;&=\; 95\% \\
P(\mu-2.58\sigma \;< X \;\leq\; \mu+2.58\sigma) \;&=\; 99\% \\
P(\mu-3.29\sigma \;< X \;\leq\; \mu+3.29\sigma) \;&=\; 99.9\%
\end{aligned}$$

Testing whether the distribution is normal for small sample set: Normality test is important in many practical problems. When the sample size is large, we can draw the histogram, see the shape, and decide, at least empirically, whether it is a normal distribution or not. This is not possible when the sample size is small (say, less than 30). A more formal way of acertaining normality is by drawing **Normal Probability Plot** as explained below.

- Step 1: Sort the data in ascending order
- Step 2: Compute $f_i = \frac{(i-0.375)}{(n+0.25)}$, where i is the index (i.e. serial number in the sorted list) of a sample and n is the total number of samples. This f_i represents the probability of finding a sample whose value is equal or less than the i^{th} sample.
- Step 3: Find the Z-score, $z_i = \frac{xi-\mu}{\sigma}$, of the i_{th} sample on the sorted list. Use the value of μ and σ calculated from the given samples. If $f_i \approx z_i$ for all i-s, the distribution is normal. Plot them, and if all the points lie on a straight line of slope 1, the sample set follows normal distribution.
- Step 4: Otherwise, as the value z_i is proportional to x_i, we can as well plot x_i on the X-axis and f_i on the Y-axis. If this plot is a straight line, we can say that the samples follow normal distribution.

Use Standard Normal Distribution Table given in the class to solve the following problems.
Examples:
1. Let X follows normal distribution with mean 0.8 and variance 4 (so that $\sigma = 2$). Find probability of $X \leq 2.44$. (X の平均値が 0.8 で分散が 4 のとき、$X \leq 2.44$ である確率を求めよ)

2. In a production of iron rods let the diameter X be normally distributed with mean 20 mm. and standard deviation 0.08 mm. (鉄の棒を製造するとき、直径 X の平均が 20mm. で標準分布し、標準偏差が 0.008 mm. のとき)

(a) What percentage of defectives can we expect if we set the tolerance limits at 20mm ± 0.2 mm. ? (直径の許容値を 2mm ± 0.2mm としたときの不良品が出来る確率を求めよ)

(b) How should we set the tolerance limits to allow for 4% defectives? (不良品を 4% まで許容するとき、許容値をいくつに設定すればいいか求めよ)

Problems:

1. Let X be normal with mean 10 and variance 4. Find $P(X > 12)$, $P(X < 10)$, $P(X < 11)$, $P(9 < X < 13)$. (平均が 10 で分散が 4 の時、$P(X > 12)$,$P(X < 10)$, $P(X < 11)$, $P(9 < X < 13)$ の X をそれぞれ求めよ)

2. Let X be normal with mean 50 and variance 9. Determine c such that $P(X < c) = 5\%$, $P(x > c) = 1\%$, $P(50 - c < X \leq 50 + c) = 50\%$ (平均が 50 で分散が 9 の時 $P(x > c) = 1\%$, $P(50 - c < X \leq 50 + c) = 50\%$ である c を求めよ)

3. If the lifetime X of a certain kind of automobile battery is normally distributed with a mean of 5 years and a standard deviation of 1 year, and the manufacturer wishes to guarantee the battery for 4 years, what percentage of the batteries will he have to replace under the guarantee? (ある種類の自動車のバッテリーの寿命 X は平均 5 年、標準偏差が 1 年で標準分布であり、メーカーがそのバッテリーの補償期間を 4 年にしている。このとき、メーカーは何%のバッテリーを補償で交換しなければならないかを求めよ)