PROBABILITY PRACTICAL 2 SOLUTIONS

DAVID STEINSALTZ

(1) Stroke patients with aphasic deficits are each given a number of straightforward tasks in a psychometric test. The number of errors made by 123 patients are shown in the table below. Calculate the mean and variance of the number of errors per patient and comment on these values. Fit a Poisson distribution and comment on how well it fits the observed data.

Number of errors	0	1	2	3	4	5 or more
Number of patients	5	30	56	15	10	7

$\bar{x} = \frac{5\times0+30\times1+56\times2+15\times3+10\times4+7\times5}{123} = 2.13$

$s^2 = \frac{5(0-\bar{x})^2+30(1-\bar{x})^2+56(2-\bar{x})^2+15(3-\bar{x})^2+10(4-\bar{x})^2+7(5-\bar{x})^2}{123-1} = 1.36$

The mean and variance are not very close which suggests a Poisson distribution may not be a good fit.

x	0	1	2	3	4	5 or more
$P(X=x)$	0.119	0.253	0.270	0.191	0.102	0.065
Expected	14.6	31.1	33.2	23.5	12.5	8.01
Observed	5	30	56	15	10	7

The fitted values are not especially good. The observed values are more peaked around the value of 2 than the fitted Poisson.

(2) Let X have uniform distribution on the interval $[a, b]$, defined as the continuous distribution whose density is constant on that interval and 0 outside it.

(a) What are the density and the cdf of this distribution?

The density is

$$f(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{if } x > b. \end{cases}$$

The cdf is

$$F(x) = \int_{-\infty}^{x} f(y)dy = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } x > b. \end{cases}$$

(b) What are the expectation and variance?

$$\begin{aligned}\mathbb{E}[X] &= \int_{-\infty}^{\infty} f(x)x\,dx \\ &= \frac{1}{b-a}\int_a^b x\,dx \\ &= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b \\ &= \frac{1}{b-a}\left[\frac{b^2-a^2}{2}\right] \\ &= \frac{b+a}{2}\end{aligned}$$

$$\begin{aligned}\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} f(x)x^2\,dx \\ &= \frac{1}{b-a}\int_a^b x^2\,dx \\ &= \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b \\ &= \frac{1}{b-a}\left[\frac{b^3-a^3}{2}\right] \\ &= \frac{b^2+ab+b^2}{3} \\ \mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \frac{b^2+ab+b^2}{3} - \frac{(b+a)^2}{4} \\ &= \frac{b^2-2ab+b^2}{12} \\ &= \frac{(b-a)^2}{12}\end{aligned}$$

(3) Suppose the mean household income in a city is £30,000, and the standard deviation is £10,000. 400 households are selected at random. Estimate the probability that the average household income in the sample is no more than £31,000. What do you think the distribution of incomes looks like? Does this matter for your answer?

Let X be the average of 400 randomly sampled incomes. By the CLT this is approximately normally distributed, with mean £30,000 and SD $£10,000/\sqrt{400} = £500$. Thus

$$Z := \frac{X - 30000}{500}$$

has approximately standard normal distribution. In other words

$$\mathbb{P}\{X \le 31000\} = \mathbb{P}\left\{\frac{x-30000}{500} \le \frac{31000-30000}{500}\right\} = \mathbb{P}\{Z \le 2\}$$

should be very close to $\Phi(2)$, where Φ is the standard normal cdf. This may be computed by `pnorm(2)` or otherwise to be 0.977.

The distribution of incomes is probably strongly right-skewed — that is, with its median and mode below its mean and a long tail of high incomes — but the CLT still implies that the mean of a large number of random incomes is still approximately normal.

(4) In a certain country the heights of adult males have mean 170cm and standard deviation 10cm, and the heights of adult females have mean 160cm and standard deviation 8cm; for each sex the distribution of heights approximates closely to a normal probability model. On the hypothesis that height is not a factor in selecting a mate, calculate the probability that

(a) a husband and wife selected at random are both taller than 164cm;

Let X =wife height, Y =husband height, and let the cdfs be F_X and F_Y. This is just the product $(1-F_X(164))(1-F_Y(164))$. We could compute it directly with the `R` command `(1-pnorm(164,170,10))*(1-pnorm(164,160,8))`, giving the result 0.224. Alternatively, we can convert this to statements about a standard normal random variable Z. For X we take $Z = (X-160)/8$. This is standard normal (mean 0, variance 1), so

$$\mathbb{P}\{X > 164\} = \mathbb{P}\{Z > 0.5\}.$$

Similarly for Y.

(b) in a randomly selected husband and wife the wife is taller than the husband;

The difference is a normal random variable. We know that the expectation of $X-Y$ is the difference in expectations -10, and the variance is the sum of the variances $8^2+10^2 = 164$, since we assumed X and Y to be independent. Thus we can compute $\mathbb{P}\{X-Y>0\}$ with `1-pnorm(0,-10,sqrt(164))`, yielding 0.217.

(c) the average height of a random couple is greater than 168cm. [You may use the fact that the sum of two independent normal random variables is also normal.]

$(X+Y)/2$ is normal with mean 165 and variance $(8^2+10^2)/4 = 41$. We than compute `1-pnorm(168,165,sqrt(41))`, yielding 0.320.