DEPARTMENT OF MATHEMATICS AND STATISTICS

53304 Outline lecture notes

1. Revision of Probability Theory

Probability distribution. There is a finite or countable *sample space* S. Probabilities p_j defined for j in S, such that $p_j \geq 0$, $\sum p_j = 1$. Then the mean ($\mathrm{E}(X)$ or μ) and variance ($\mathrm{V}(X)$ or σ^2) are defined by

$$\mathrm{E}(X) = \mu = \sum jp_j \quad ; \quad \mathrm{V}(X) = \mathrm{E}(X-\mu)^2 = \mathrm{E}(X^2) - \mu^2.$$

Example 1. Rolling a fair die. Uniform distribution on $S = \{1, \ldots, 6\}$, $p_j = \frac{1}{6}$ for j in S. The mean and variance are 3.5 and 35/12.

More generally, for the uniform distribution on $\{1, \ldots, n\}$, $\mu = \frac{n+1}{2}$, $\sigma^2 = \frac{n^2-1}{12}$, and the s.d. is close to $n/\sqrt{12}$ for large n.

Exercise 1. Let X, Y, Z be the results of three independent rolls of a fair die. Find the distribution of $X+Y$, and $X+Y+Z$. (NB. It is trivial to find the means (7, 10.5) and variances (35/6, 35/4)! These are both symmetrical distributions.)

Maximum and minimum of probability distributions.

Let X, Y denote two independent rolls of a fair die. Find the probability that both are no greater than 4 , i.e. that $Z \equiv \max(X,Y) \leq 4$.

Now $Z \leq 4$ iff $X \leq 4$ & $Y \leq 4$. The probability is $F_X(4)F_Y(4)$, where F_X, F_Y denote the cumulative d.f.'s of X, Y. More generally $F_Z(j) = F_X(j)F_Y(j)$.

Exercise 2. Find the c.d.f. of $\min(X,Y)$. Start by considering the event $\{\min(X,Y) \geq j\}$. [$F_X + F_Y - F_X F_Y$.]

Example 2. There are M pupils in an S5 Geography class now. The probability that a pupil takes SYS Geography is q. The probability that a pupil in SYS Geography will take part in a field trip is p. Assuming that the pupils are independent, show that the distribution of the number X of the pupils who will take part in the excursion has a Binomial distribution with parameters M, pq.

Note that, if we think of the chance of a pupil's being in the SYS Geography class and wishing to go on the excursion as pq, this is obvious. However a more detailed calculation follows. Denote by N the number of the current S5 pupils who will take SYS Geography. This follows a Binomial distribution with parameters M, q. Given that there will be N pupils in SYS Geography, the number who will go on the excursion follows the Binomial distribution with parameters N, p.

$$
\begin{aligned}
\mathrm{P}(X=k) &= \sum_{j=k}^{M} \mathrm{P}(N=j \quad \& \quad X=k) = \sum_{j=k}^{M} \mathrm{P}(N=j)\mathrm{P}(X=k \mid N=j) \\
&= \sum_{j=k}^{M} \binom{M}{j} q^j (1-q)^{M-j} \binom{j}{k} p^k (1-p)^{j-k} \\
&= p^k q^k \binom{M}{k} \sum_{s=0}^{M-k} \binom{M-k}{s} q^s (1-p)^s (1-q)^{M-k-s} \quad (s \text{ denoting } j-k) \\
&= \binom{M}{k} (pq)^k (1-pq)^{M-k}.
\end{aligned}
$$

This gives the distribution of the sum of a binomially distributed number of independent Bernoulli variables.

Geometric distribution. This models the following situation. Repeat independent 'experiments' independently until the first 'success' is obtained. How many experiments have been carried out? It is assumed that the probability p of success is the same for each experiment. Here the sample space is infinite. It consists of the integers from 1 upwards. (Note that when you use this distribution 'experiment' and 'success' are not necessarily the appropriate words to use. It depends on the context.)

Here $p_j = (1-p)^{j-1}p$. The probabilities are in geometric progression. Intuitively it is not surprising that the mean is $1/p$. (If the chance of success is p each time, it takes on average $1/p$ attempts to get one success.) To get the variance the best way is to use the probability generating function (pgf). Recall that for a discrete probability distribution X, the pgf G_X is defined by $G_X(s) = \mathrm{E}(s^X) = \sum p_j s^j$. For the Geometric distribution we get

$$
\sum_{j=1}^{\infty} (1-p)^{j-1} p s^j = ps \sum ((1-p)s)^{j-1} = \frac{ps}{(1-s+ps)}.
$$

The mean $\mu \equiv G'(1) = 1/p$,
and $\mathrm{E}(X(X-1)) \equiv G''(1)$ (see 53201), so $\sigma^2 \equiv \mathrm{V}(X) = G''(1) + \mu - \mu^2 = (1-p)/p^2$.

Another useful fact to recall about the Geometric distribution is that $\mathrm{P}(X > k)$ is the probability that the first k attempts are not successes, which is $(1-p)^k$.

Poisson distribution. This is used to model the 'random' distribution of objects in space or time. E.g. the number of vehicles in 100-metre stretches of a motorway, the number of customers arriving at a shop in periods of 5 minutes, the number of seeds landing on square metres of ground, the number of cells in 1 ml of a suspension. The positive real parameter λ is used to denote the mean of the distribution. Again the sample space is infinite. The values X usually represent *counts* and are thus integers from 0 upwards. The probability

$$p_j = e^{-\lambda}\frac{\lambda^j}{j!}, \quad (j = 0, \ldots), \qquad \mathrm{E(X)} = \lambda, \qquad \mathrm{V(X)} = \lambda.$$

You should note that here $e^{-\lambda}$ is just a multiplicative constant. The part that determines the shape of the distribution is $\lambda^j/j!$.

Exercise 3. Calculate Poisson probabilities to 3 d.p. for $\lambda = 1,\ 3,\ 1.5,\ 0.8$.

[Working for $\lambda = 0.8$.

j	0	1	2	3	4	5
P	.449	.359	.144	.038	.008	.001

The 0th entry is $e^{-0.8}$. Then to get the jth number from the previous one, multiply by $0.8/j$.]

Which probability is greatest? I.e. what is the *mode* of the distribution?

Example 3. Let X, Y denote independent Poisson distributions with means λ, μ. Find the distribution of $X + Y$.

$$\mathrm{P}(X+Y=j) = \sum_{k=0}^{j}\mathrm{P}(X=k \ \&\ Y=j-k) = \frac{e^{-(\lambda+\mu)}}{j!}\sum_{k=0}^{j}\binom{j}{k}\lambda^k\mu^{j-k} = e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^j}{j!}.$$

This is the Poisson distribution with mean $\lambda+\mu$. You can think of λ as the rate of the passing of cars, and μ as the rate of other vehicles. Then $\lambda+\mu$ represents the rate of passing of all vehicles.

Probability generating function

If a discrete random variable X has the probability distriubtion

$$\mathrm{P}(X = i) = p_i, \quad i = 0, 1, 2, \cdots, r,$$

then the probability generating function (p.g.f.) of the r.v. is defined by

$$G(s) = G_X(s) = \mathrm{E}(s^X) = p_0 + p_1 s + p_2 s^2 + \cdots p_r s^r.$$

Conversely, if X has its p.g.f.

$$G(s) = G_X(s) = \mathrm{E}(s^X) = p_0 + p_1 s + p_2 s^2 + \cdots p_k s^k.$$

then its probability distribution is

$$\mathrm{P}(X = i) = p_i, \quad i = 0, 1, 2, \cdots, k.$$

It is useful to know that

$$\mathrm{E}(X) = G'(1),$$

where $G'(s) = dG(s)/ds$.

Exponential distribution and its memoryless property

A continuous r.v. W is said to follow an exponential distribution with parameter (or rate) μ if it has the p.d.f.

$$f(w) = \begin{cases} \mu e^{-\mu w} & \text{if } w \geq 0, \\ 0 & \text{if } w < 0. \end{cases}$$

It can be shown that

$$\mathrm{E}(W) = \frac{1}{\mu}$$

and

$$\mathrm{P}(W > c) = e^{-\mu c}, \quad \forall c > 0.$$

It can also be shown that

$$\mathrm{P}(W > c + h | W > c) = e^{-\mu h}, \quad \forall c, h > 0,$$

which is called the memory less property.