2. Discrete random variables

If the chance outcome of the experiment is a **number**, it is called a *random variable*.

Discrete random variable: the possible outcomes can be listed e.g. 0, 1, 2, ..., or yes/no, or A, B, C.. etc.

Notation for random variables: capital letter near the end of the alphabet e.g. X, Y.

P(X = k) denotes "The probability that X takes the value k".

Note: $0 \le P(X = k) \le 1$ for all k, and $\sum_k P(X = k) = 1$.

Mean (or expected value) of a variable

For a random variable X taking values 0, 1, 2, ... , the mean value of X is:

$$\mu = \sum_k k\, P(X = k) = 0 \times P(X = 0) + 1 \times P(X = 1) + 2 \times P(X = 2) + \cdots$$

The mean is also called: population mean
expected value of X, (or E(X) or $\langle X \rangle$)
average of X

Intuitive idea: if X is observed in repeated independent experiments and $\bar{X}_n = \frac{1}{n}\sum_{i+1}^{n} X_i$ is the sample mean after n observations then as n gets bigger, $\bar{X}_n$ tends to μ.

Mean (or expected value) of a function of a random variable

For a random variable X, the expected value of $f(X)$ is given by

$$E\big(f(X)\big) \equiv \langle f(X) \rangle = \sum_k f(k) P(X = k)$$

For example $f(X)$ might give the winnings on a bet on X. The expected winnings of a bet would be the sum of the winnings for each outcome multiplied by the probability of each outcome.

Note: “expected value” is a somewhat misleading technical term, equivalent in normal language to the mean or average of a population. The expected value is not necessarily a likely outcome, in fact often it is impossible. It is the average you would expect if the variable were sampled very many times.

E.g. you bet on a coin toss: Heads (H) wins 50p, Tails (T) loses £1. The expected “winnings” W is

$$E(W) = \langle W \rangle = £0.5 \times P(H) + (-£1) \times P(T) = £0.5 \times \frac{1}{2} - £1 \times \frac{1}{2} = -£0.25$$

Sums of expected values

Means simply add, so e.g.

$$\langle f(X)+g(X)\rangle = \sum_k \big(f(X)+g(X)\big)P(X=k) = \sum_k f(X)P(X=k) + \sum_k g(X)P(X=k)$$
$$= \langle f(X)\rangle + \langle g(X)\rangle$$

This also works for functions of two (or more) different random variables X and Y.

Variance and standard deviation of a distribution

For a random variable X taking values 0, 1, 2 the mean μ is a measure of the average value of a distribution, $\mu = \langle k\rangle$.

The standard deviation, σ, is a measure of how spread out the distribution is. The variance is the mean-squared deviation from the mean, given by σ^2. So σ is the r.m.s.

Variance

$$\sigma^2 \equiv \mathrm{var}(X) = \langle (k-\mu)^2\rangle = \sum_k (k-\mu)^2 P(X=k)$$

Note that

$$\langle (k-\mu)^2\rangle = \langle k^2 - 2\,k\,\mu + \mu^2\rangle = \langle k^2\rangle - 2\,\langle k\,\rangle\mu + \mu^2$$
$$= \langle k^2\rangle - 2\mu^2 + \mu^2 = \langle k^2\rangle - \mu^2$$

So the variance can also be written

$$\sigma^2 = \mathrm{var}(X) = \langle k^2\rangle - \mu^2 = \langle k^2\rangle - \langle k\rangle^2 = \sum_k k^2 P(X=k) - \mu^2$$

This equivalent form is often easier to evaluate in practice, though can be less numerically stable (e.g. when subtracting two large numbers).

Sums of variances

For two <u>independent</u> (or just uncorrelated) random variables X and Y the variance of X+Y is given by the sums of the separate variances:

If X has $\langle X\rangle = A$, and Y has $\langle Y\rangle = B$, then $\langle X+Y\rangle = \langle X\rangle + \langle Y\,\rangle = A+B$. Hence

$$\mathrm{var}(X+Y) = \langle (X+Y-A-B)^2\rangle = \langle [(X-A)+(Y-B)]^2\rangle$$
$$= \langle (X-A)^2\rangle + \langle (Y-B)^2\rangle + 2\langle (X-A)(Y-B)\rangle$$

If X and Y are independent (or just uncorrelated) then $\langle (X-A)(Y-B)\rangle = \langle (X-A)\rangle\langle (Y-B)\rangle = 0$ hence

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y).$$

The generalizes easily so that the variance of a sum of any number of independent variables is given by the sum of the variances.

Binomial distribution

A process with two possible outcomes, "success" and "failure" (or yes/no, etc) is called a *Bernoulli trial.*

e.g. coin tossing: Heads or Tails
quality control: Quality Satisfactory or Unsatisfactory

> An experiment consists of n independent Bernoulli trials and p = probability of success for each trial. Let X = total number of successes in the n trials.
> Then $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $k = 0, 1, 2, \ldots, n$.

This is called the Binomial distribution with parameters n and p, or $B(n, p)$ for short.

$X \sim B(n, p)$ stands for "X has the Binomial distribution with parameters n and p."

Justification

"$X = k$" means k successes (each with probability p) and n-k failures (each with probability 1-p). Suppose for the moment all the successes come first:

$$\text{probability} = p \times p \times p \ldots \times p \times (1-p) \times (1-p) \times \ldots \times (1-p) \text{ (by independence)}$$

$$= p^k (1-p)^{n-k}$$

Every possible different ordering also has this same probability. The total number of ways of choosing k out of the n trails to be successes is $\binom{n}{k}$, so there are $\binom{n}{k}$, possible orderings. Since each ordering is an exclusive possibility, by the special addition rule the overall probability is

$$\binom{n}{k} p^k (1-p)^{n-k}$$

Situations where a Binomial might occur

1) Quality control: select n items at random; X = number found to be satisfactory.

2) Survey of n people about products A and B; X = number preferring A.

3) Telecommunications: n messages; X = number with an invalid address.

4) Number of items with some property above a threshold; eg. X = number with height > A

***Example**: A component has a 20% chance of being a dud. If five are selected, what is the probability that more than one is a dud?*

Solution

Let X = number of duds in selection of 5

Bernoulli trial: dud or not dud, X ~ B(5, 0.2)

P(More than one dud) = $P(X > 1) = 1 - P(X \leq 1)$
$= 1 - C_0^5 0.2^0(1-0.2)^5 - C_1^5 0.2^1(1-0.2)^4$
$= 1 - 1 \times 1 \times 0.8^5 - 5 \times 0.2 \times 0.8^4$
= 1 - 0.32768 - 0.4096 ≈ 0.263.

Mean and variance of a binomial distribution

If $X \sim B(n,p)$, then $\mu = E(X) = np$, and σ^2 = var(X) = $np(1-p)$

Derivation

Suppose first that we have a single Bernoulli trial. Assign the value 1 to success, and 0 to failure, the first occurring with probability p and the second having probability $1 - p$. The expected value in this trial will be equal to $\mu = 1 \times p + 0 \times (1-p) = p$. Since the *n* trials are independent, the total expected value is just the sum of the expected values for each trial, hence *np*.

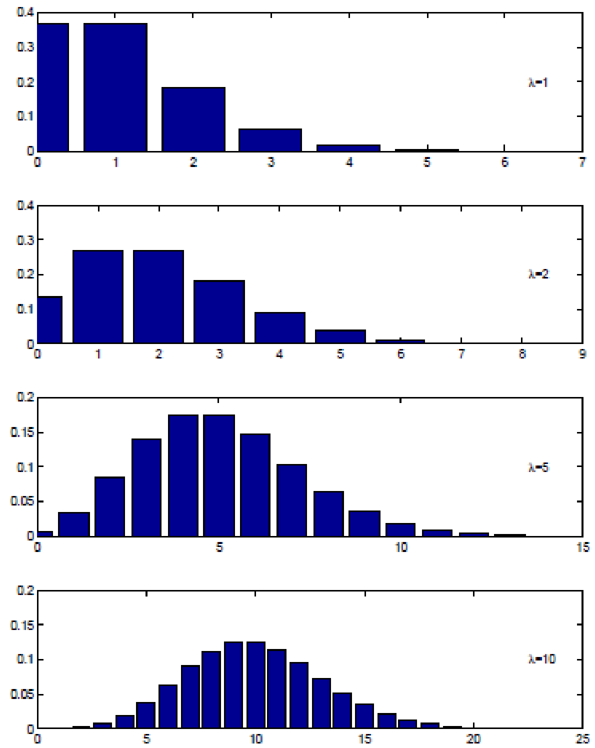
The variance in a single trial is:

$$\sigma^2 = \langle X^2 \rangle - \langle X \rangle^2 = 1^2 \times p + 0^2 \times (1-p) - p^2 = p(1-p)$$

Hence the variance for the sum of *n* independent trials is, by the rule for summing the variances of independent variables, $np(1-p)$.

Poisson distribution

If events happen independently of each other, with average number of events in some fixed interval λ, then the distribution of the number of events k in that interval is Poisson:



A random variable Y has the Poisson distribution with parameter λ (> 0) if:

$$P(Y = k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad (k = 0,1,2, \ldots)$$

Justification (non-examinable): Consider splitting the time T into lots of N short intervals, each of duration $\frac{T}{N}$. Assuming the event probability is constant in time, for large N the probability to have an event in one interval is then $\frac{\lambda}{N}$ (which becomes very small), and the probability of getting k intervals with events (and $N-k$ non-events) is then given by the Binomial distribution

$$\mathrm{P(k)} = \binom{N}{k}\left(\frac{\lambda}{N}\right)^k\left(1-\frac{\lambda}{N}\right)^{N-k}$$

Taking the limit of large N, $N \gg k$, as we divide into smaller and smaller time intervals λ/N becomes very small and k can be neglected compared to N, so

$$\ln\left(1-\frac{\lambda}{N}\right)^{N-k} = (N-k)\ln\left(1-\frac{\lambda}{N}\right) \approx N\ln\left(1-\frac{\lambda}{N}\right) \approx -\lambda \Rightarrow \left(1-\frac{\lambda}{N}\right)^{N-k} \approx e^{-\lambda}$$

(where we used the series expansion $\ln(1-x) \approx -x - x^2/2 + \cdots \approx -x$ for small x). Hence

$$\mathrm{P(k)} \approx \frac{\lambda^k e^{-\lambda}}{k!}\frac{N!}{N^k(N-k)!} \approx \frac{\lambda^k e^{-\lambda}}{k!}\frac{N(N-1)\ldots(N-k+1)}{N^k} \approx \frac{\lambda^k e^{-\lambda}}{k!}$$

Mean and variance

By construction the mean of a Poission distribution is λ. For the variance we know the result for the Binomial distribution is $np(1-p)$. The Poisson distribution corresponds to the limit of p small and n large, so $1-p \approx 1$. Hence the variance in this limit is np, which is the same as the mean. So a Poisson distribution has $\mathrm{var}(X) = E(X) = \lambda$.

If X ~ Poisson with mean λ, then $\mu = E(X) = \lambda$, and $\sigma^2 = \mathrm{var}(X) = \lambda$.

Examples of possible Poisson processes

1) Number of messages arriving at a telecommunications system in a day

2) Number of flaws in a fibre optic cable.

3) Number of radio-active particles detected in a given time

4) Number of photons arriving at a CCD pixel in some exposure time
   (e.g. astronomy observations)

Sum of Poisson variables

If X is Poisson with average number λ_X and Y is Poisson with average number λ_Y, then the total number $X+Y$ is Poisson with average number $\lambda_X + \lambda_Y$. In particular the probability of events per unit time does not need to be constant for the total to be Poisson: you can split up the total into a sum of the number of events in smaller intervals. In the example of the messages arriving at the telecommunications system most of the message might arrive during the day, with many fewer at night. However the total number of messages in a day is still Poisson, because the number during the day and the number during the night both are.

Example: Telecommunications.

Messages arrive at a switching centre at random and at an average rate of 1.2 per second.

(a) Find the probability of 5 messages arriving in a 2-sec interval.

(b) For how long can the operation of the centre be interrupted, if the probability of losing one or more messages is to be no more than 0.05?

Solution

Times of arrivals form a Poisson process, rate $\nu = 1.2/\mathrm{sec}$.

(a) Let Y = number of messages arriving in a 2-sec interval.

Then Y ~ Poisson, mean number $\lambda = \nu t = 1.2 \times 2 = 2.4$. Hence

$$P(Y=5) = \frac{e^{-2.4}2.4^5}{5!} = 0.060$$

(b) Let the required time = t seconds.

Let W = number of messages in t seconds, so that $W \sim$ Poisson,$\lambda = 1.2 \times t = 1.2t$

P(At least one message) = $P(W \geq 1) = 1 - P(W = 0) = 1 - e^{-1.2t} \leq 0.05.$

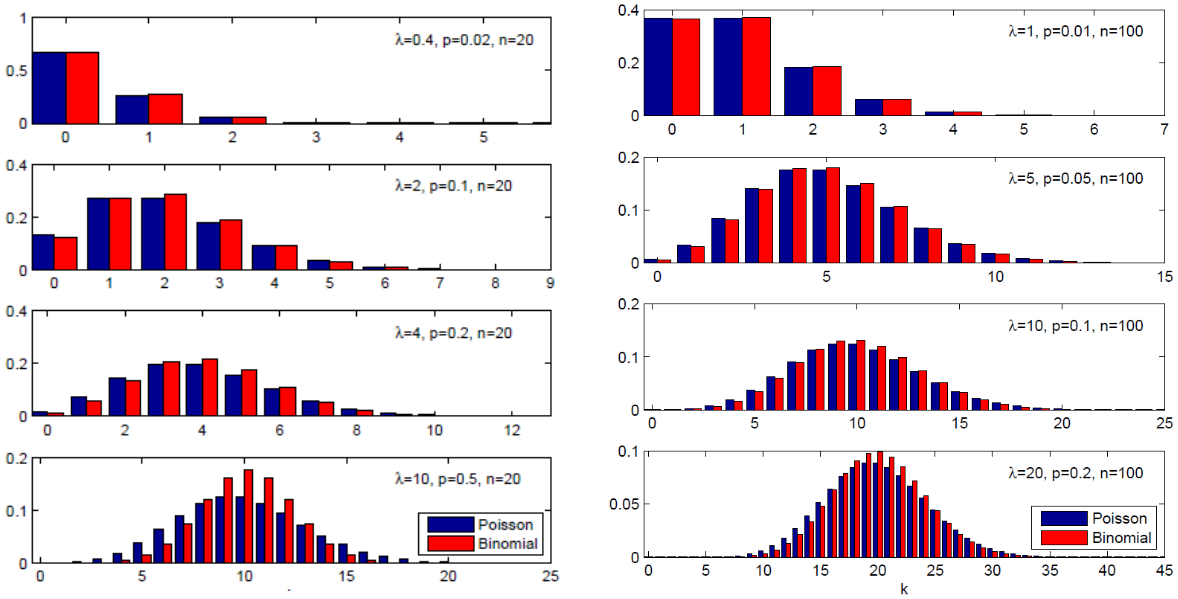
$$\Rightarrow e^{-1.2t} \geq 0.95$$

$$\Rightarrow -1.2t \geq \ln(0.95) = -0.05129$$

$$\Rightarrow t \leq 0.043 \text{ seconds}$$

Approximation to the Binomial distribution

The Poisson distribution is an approximation to B(n, p), when n is large and p is small (e.g. if $np < 7$, say); in this case, if $X \sim B(n, p)$ then $P(X = k) \approx \frac{e^{-\lambda}\lambda^k}{k!}$ Where $\lambda = np$ i.e. X is approximately Poisson, parameter np.



Example: The probability of a certain part failing within ten years is 10^{-6}. Five million of the parts have been sold so far; what is the probability that three or more will fail within ten years?

Solution

Let X = number failing in ten years, out of 5,000,000.

$X \sim$ B(5000000, 10^{-6})

Evaluating the Binomial probabilities is rather awkward; better to use the Poisson approximation.

X has approximately Poisson distribution with $\lambda = np = 5000000 \times 10^{-6} = 5.$

∴ P(Three or more fail) $= P(X \geq 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$

$$= 1 - \frac{e^{-5}5^0}{0!} - \frac{e^{-5}5^1}{1!} - \frac{e^{-5}5^2}{2!}$$
$$= 1 - e^{-5}(1 + 5 + 12.5) = 0.875$$

For such small p and large n the Poisson approximation is very accurate (exact result is also 0.875 to three significant figures).