1 Discrete Probability Models

We have seen how to relate events to sets, and how to calculate probabilities for events by working with the sets that represent them. *Discrete probability models* provide a framework for thinking about discrete *random quantities*, and *continuous probability models* form a frame- work for thinking about *continuous random quantities.*

Example. Consider the sample space for tossing a fair coin twice:

$$S = \{HH, HT, TH, TT\}.$$

These outcomes are equally likely. There are several random variable/quantities we could associate with this experiment. For example, we could count the number of heads, or the number of tails.

Formally, a *random variable* is a real valued function which acts on *elements* of the sample space (outcomes). That is, to each outcome, the *random variable* assigns a real number. Random variables are always denoted by upper case letters. In our example, if we let X be the number of heads, we have

$$\begin{aligned} X(HH) &= 2 \\ X(HT) &= 1 \\ X(TH) &= 1 \\ X(TT) &= 0. \end{aligned}$$

The observed value of a random quantity is the number corresponding to the actual outcome. That is, if the outcome of an experiment is $s \in S$, then $X(s) \in \mathbb{R}$ is the observed value. This observed value is always denoted with a lower case letter — here x. Thus $X = x$ means that the observed value of the random quantity, X is the number x. The set of possible observed values for X is

$$S_X = \{X(s)|s \in S\}.$$

For the above example we have $S_X = \{0, 1, 2\}$. Clearly here the values are not all equally likely.

Example. Roll one die and call the random number which is uppermost Y. The sample space for the *random variable* Y is $S_Y = \{1, 2, 3, 4, 5, 6\}$ and these outcomes are all equally likely. Now roll two dice and call their sum Z. The sample space for Z is $S_Z =$

$\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ and these outcomes are *not* equally likely. However, we know the probabilities of the events corresponding to each of these outcomes, and we could display them in a table as follows.

Outcome	2	3	4	5	6	7	8	9	10	11	12
Probability	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

This is essentially a tabulation of the *probability mass function* for the random quantity Z.

1.1 Probability mass functions (PMFs)

For any discrete random variable X, we define the *probability mass function* (PMF) to be the function which gives the probability of each $x \in S_X$. Clearly we have

$$P(X = x) = \sum_{\{x \in S | X(s) = x\}} P(\{s\}).$$

That is, the probability of getting a particular number is the sum of the probabilities of all those outcomes which have that number associated with them. Also $P(X = x) \geq 0$ for each $x \in S_X$, and $P(X = x) = 0$ otherwise. The set of all pairs $\{(x, P(X = x)) | x \in S_X\}$ is known as the *probability distribution* of X.

Example. For the example above concerning the sum of two dice, the probability distribution is

$$\{(2, 1/36), (3, 2/36), (4, 3/36), (5, 4/36), (6, 5/36), (7, 6/36), (8, 5/36), (9, 4/36), (10, 3/36), (11, 2/36), (12, 1/$$

and the probability mass function can be tabulated as

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

1.2 Cumulative distribution functions (CDFs)

For any discrete random quantity, X, we clearly have

$$\sum_{\{x \in S_X\}} P(X = x) = 1$$

as every outcome has some number associated with it. It can often be useful to know the probability that your random number is no greater than some particular value. With that in mind, we define the *cumulative distribution function*,

$$F_X(x) = P(X \leq x) = \sum_{\{y \in S_X | y \leq x\}} P(X = y).$$

Example. For the sum of two dice, the CDF can be tabulated for the outcomes as

x	2	3	4	5	6	7	8	9	10	11	12
$F_X(x)$	1/36	3/36	6/36	10/36	15/36	21/36	26/36	30/36	33/36	35/36	36/36

but it is important to note that the CDF is defined for all real numbers — not just the possible values. In our example we have

$$\begin{aligned} F_X(-3) &= P(X \leq -3) = 0, \\ F_X(4.5) &= P(X \leq 4.5) = P(X \leq 4) = 6/36, \\ F_X(25) &= P(X \leq 25) = 1. \end{aligned}$$

1.3 Expectation and variance for discrete random quantities

1.3.1 Expectation

Just as it is useful to summarize data, it is just as useful to be able to summarize the distribution of random quantities. The *location measure* used to summarize random quantities is known as the *expectation* of the random quantity. It is the *"center of mass"* of the probability distribution. The expectation of a discrete random quantity X, written $E(X)$ is defined by

$$E(X) = \sum_{x \in S_X} x P(X = x).$$

The expectation is often denoted by μ_X or even just μ. Note that the expectation is a known function of the probability distribution. It is *not* a random quantity, and in particular, it is *not* the sample mean of a set of data (random or otherwise). In fact, there is a *relationship* between the sample mean of a set of data and the expectation of the underlying probability distribution generating the data.

Example. For the sum of two dice, X, we have

$$E(X) \quad = \quad 2 \times \frac{1}{36} + 3 \times \frac{2}{36} + 4 \times \frac{3}{36} + ... + 12 \times \frac{1}{36} = 7.$$

By looking at the symmetry of the mass function, it is clear that in some sense 7 is the "central" value of the probability distribution.

1.3.2 Variance

We now have a method for summarizing the location of a given probability distribution, but we also need a summary for the *spread*. For a discrete random quantity X, the *variance* of X is defined by

$$V(X) = \sum_{x \in S_X} \left[(x\text{-}E(X))^2 P(X = x) \right].$$

The variance is often denoted σ_X^2 , or even just σ^2. Again, this is a known function of the probability distribution. It is not random, and it is not the *sample* variance of a set of data. Again, the two are related. The variance can be re-written as

$$V(X) = \sum_{x_i \in S_X} x_i^2 P(X = x_i) - [E(X)]^2.$$

and this expression is usually a bit easier to work with. We also define the *standard deviation* of a random variable by

$$SD(X) = \sqrt{V(X)},$$

and this is usually denoted by σ_X or just σ.

Example. For the sum of two dice, X, we have

$$\sum_{x_i \in S_X} x_i^2 P(X = x_i) \quad = \quad 2^2 \times \frac{1}{36} + 3^2 \times \frac{2}{36} + 4^2 \times \frac{3}{36} + ... + 12^2 \times \frac{1}{36} = \frac{329}{6}.$$

and so

$$V(X) = \frac{329}{6} - 7^2 = \frac{35}{6}.$$

1.3.3 Properties of expectation and variance

One of the reasons that expectation is widely used as a measure of location for probability distri- butions is the fact that it has many desirable mathematical properties which make it elegant and convenient to work with. Indeed, many of the nice properties of expectation lead to corresponding nice properties for variance, which is one of the reasons why variance is widely used as a measure of spread.

Expectation of a function of a random variable

Suppose that X is a discrete random variable, and that Y is another random variable that is a known function of X. That is, $Y = g(X)$ for some function g. What is the expectation of Y?

Example. Throw a die, and let X be the number showing. We have

$$S_X = \{1, 2, 3, 4, 5, 6\}$$

and each value is equally likely. Now suppose that we are actually interested in the square of the number showing. Define a new random quantity $Y = X^2$. Then

$$S_Y = \{1, 4, 9, 16, 25, 36\}$$

and clearly each of these values is equally likely. We therefore have

$$E(Y) \quad = \quad 1 \times \frac{1}{6} + 4 \times \frac{1}{6} + 9 \times \frac{1}{6} + ... + 36 \times \frac{1}{6} = \frac{91}{6}.$$

The above example illustrates the more general result, that for $Y = g(X)$, we have

$$E(Y) = \sum_{x \in S_X} g(x) P(X = x).$$

Note that in general $E(g(X)) \neq g(E(X))$. For the above example, $E\left(X^2\right) = 91/6 \simeq 15.2$, and $[E(X)]^2 = 3.5^2 = 12.25$. We can use this more general notion of expectation in order to redefine variance purely in terms of expectation as follows:

$$V(X) = E[X - E(X)]^2 = E\left(X^2\right) - E(X)^2.$$

Having said that $E(g(X)) \neq g(E(X))$ in general, it does in fact hold in the (very) special, but important case where g is a linear function.

Properties

Let X, Y be random variables and a, b are are known real constants, then we have

1. $E(aX + b) = aE(X) + b$
2. $E(X + Y) = E(X) + E(Y)$
3. $E(XY) = E(X)E(Y)$; where X and Y are independent random variables
4. $V(aX + b) = a^2V(X)$
5. $V(X + Y) = V(X) + V(Y)$; where X and Y are independent random variables.

Proof. 1.

$$\begin{aligned}
E(aX + b) &= \sum_{x \in S_X} (ax + b)P(X = x) \\
&= \sum_{x \in S_X} axP(X = x) + \sum_{x \in S_X} bP(X = x) \\
&= a\sum_{x \in S_X} xP(X = x) + b\sum_{x \in S_X} P(X = x) \\
&= aE(X) + b.
\end{aligned}$$

4.

$$\begin{aligned}
V(aX + b) &= E[(aX + b) - E(aX + b)]^2 = E\left(X^2\right) - E(X)^2 \\
&= E[aX + b - (aE(X) + b)]^2 \\
&= E[aX - aE(X)]^2 \\
&= a^2E[X - E(X)]^2 \\
&= a^2V(X).
\end{aligned}$$

5.

$$\begin{aligned}
Var(X + Y) &= E[X + Y]^2 - [E(X + Y)]^2 \\
&= E\left(X^2 + 2XY + Y^2\right) - [E(X) + E(Y)]^2 \\
&= E\left(X^2\right) + 2E(XY) + E\left(Y^2\right) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \\
&= E\left(X^2\right) + 2E(X)E(Y) + E\left(Y^2\right) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \\
&= E\left(X^2\right) - E(X)^2 + E\left(Y^2\right) - E(Y)^2 \\
&= V(X) + V(Y).
\end{aligned}$$

Proof of 2. and 3. are omitted here. □

1.4 Bernoulli distribution

Suppose that we have an event E in which we are interested, and we write its sample space as

$$S = \{E, E^c\}.$$

We can associate a random variable with this sample space, traditionally denoted I, as $I(E) = 1$, $I(E^c) = 0$. So, if $P(E) = p$, we have

$$S_I = \{0, 1\},$$

and$P(I = 1) = p$, $P(I = 0) = 1 - p$. This random quantity,I is known as an *indicator variable*, and is often useful for constructing more complex random quantities. We write

$$I \sim Bern(p).$$

We can calculate its expectation and variance as follows.

$$\begin{aligned} E(I) &= 0 \times (1-p) + 1 \times p = p \\ E(I^2) &= 0^2 \times (1-p) + 1^2 \times p = p \\ V(I) &= E\left(I^2\right) - E(I)^2 \\ &= p\text{-}p^2 = p(1-p) \end{aligned}$$

With these results, we can now go on to understand the binomial distribution.

1.5 The binomial distribution

The binomial distribution is the distribution of the number of "successes" in a series of n independent "trials", each of which results in a "success" (with probability p) or a "failure" (with probability $1 - p$). If the number of successes is X, we would write

$$X \sim B(n, p)$$

to indicate that X is a binomial random quantity based on n independent trials, each occurring with probability p.

Example.

1. Toss a fair coin 100 times and let X be the number of heads. Then $X \sim B(100, 0.5)$.
2. A certain kind of lizard lays 8 eggs, each of which will hatch independently with probability 0.7. Let Y denote the number of eggs which hatch. Then $Y \sim B(8, 0.7)$.

Let us now derive the probability mass function for $X \sim B(n,p)$. Clearly X can take on any value from 0 up to n, and no other. Therefore, we simply have to calculate $P(X = x)$ for $x = 0, 1, 2, ..., n$. The probability of x successes followed by $n - x$ failures is clearly $p^x(1-p)^{n-x}$. Indeed, this is the probability of *any* particular sequence involving x successes. There are $\binom{n}{x}$ such sequences, so by the multiplication principle, we have

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, 2, ..., n.$$

Properties of binomial distribution

It is possible (but a little messy) to derive the expectation and variance of the binomial distribution directly from the PMF. However, we can deduce them rather more elegantly if we recognize the relationship between the binomial and Bernoulli distributions. If $X \sim B(n,p)$ then

- $E(X) = np$
- $V(X) = np(1-p)$.

Exercise. Observation over a long period of time has shown that a particular sales man can make a sale on a single contact with the probability of 20%. Suppose the same person contact four prospects,

1. What is the probability that exactly 2 prospects purchase the product?
2. What is the probability that at least 2 prospects purchase the product?
3. What is the probability that all the prospects purchase the product?
4. What is the expected value of the prospects that would purchase the product?

Solution. Let X denote the number of prospect: $x = 0, 1, 2, 3, 4$. Let p denote the probability of (success) purchase = 0.2 Hence, X $B(4, 0.2)$.

1. $P(X = 2) = \binom{4}{2} 0.2^2 0.8^{4-2} = 0.1536$
2. $P(X \geq 2) = P(2) + P(3) + P(4) = 0.1536 + \binom{4}{3} 0.2^3 0.8^{4-3} + \binom{4}{4} 0.2^4 0.8^{4-4} = 0.1536 + 0.0256 + 0.0016 = 0.1808$
3. $P(X = 4) = \binom{4}{4} 0.2^4 0.8^{4-4} = 0.0016$
4. Expected value $= E(X) = np = 4 \times 0.2 = 0.8$

1.6 The geometric distribution

Consider a random experiment in which all the conditions of a binomial distribution hold. However, instead of fixed number of trials, trials are conducted until first success occurs. Hence by definition, in a series of independent binomial trials with constant probability p of success, let the random variable X denotes number of trials until first success. Then X is said to have a geometric distribution, denoted by

$$X \sim Geom(p),$$

with parameter p and given by

$$P(X = x) = (1-p)^{x-1} p, \qquad x = 1, 2, ...$$

Example. If the probability that a wave contain a large particles of contamination is 0.01, it assumes that the wave are independent, what is the probability that exactly 125 waves need to be analyzed before a large particle is detected?

Solution. Let X denotes the number of samples analyzed until a large particle is detected. Then $X \sim Geom(p)$ is a geometric random variable with $p = 0.01$. Hence, the required probability is $P(x = 125) = (0.01)\,(0.99)^{124} = 0.0029$.

Properties of geometric distribution

Suppose that $X \sim Geom(p)$. Then the mean and variance,

- $E(X) = \frac{1}{p}$
- $V(X) = \frac{1-p}{p^2}$.

1.7 The Poisson distribution

The Poisson distribution is a very important discrete probability distribution, which arises in many different contexts in probability and statistics. Typically, Poisson random quantities are used in place of binomial random quantities in situations where the sample n is large, the probability of obtaining success in any one trial p is small, and the expectation np is stable.

Example. Consider the number of calls made in a 1 minute interval to an Internet service provider (ISP). The ISP has thousands of subscribers, but each one will call with a very small probability. The ISP knows that on average 5 calls will be made in the interval. The actual number of calls will be a Poisson random variable, with mean 5.

A Poisson random variable, X with parameter λ is denote

$$X \sim Poi(\lambda),$$

and it's distribution is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, ...$$

Example. Flaws occur at random along the length of a thick, suppose that a number of flaws follows a Poisson distribution with a mean flaw of 2.3 per mm. determine probability of exactly 2 flaws in one mm of wire.

Solution. $P(X = 2) = \frac{e^{-2.3}2.3^2}{2!} = 0.265$.

The Poisson distribution has an approximation to binomial; when n is large and p is as close to 0 as possible, then the Poisson distribution has a history which approximate of that of the binomial. we can apply the Poisson distribution to the binomial when $n \geq 30$ and $np < 5$.

Exercise. If the 3% of the electric doors manufactured by a company are defective. Find the probability that in the sample of 120 doors, at most 3 doors are defective.

1. Use binomial to solve the problem.

2. Use Poisson distribution and compare your results.

Properties of The Poisson distribution

Suppose that $X \sim Poi(\lambda)$. Then the mean and variance,

- $E(X) = V(X) = \lambda$.

2 Continuous Probability Models

We now a have a good understanding of discrete probability models, but we haven't developed any techniques for handling continuous random variables. These are random variables with a sample space which is neither finite nor countably infinite. The sample space is usually taken to be the real line, or a part thereof. Continuous probability models are appropriate if the result of an experiment is a continuous *measurement*, rather than a *count* of a discrete set.

If X is a continuous random variable with sample space S_X, then for any particular $a \in S_X$, we generally have that

$$P(X = a) = 0.$$

This is because the sample space is so "large" and every possible outcome so "small" that the probability of any *particular* value is vanishingly small. Therefore the probability mass function we defined for discrete random quantities is inappropriate for understanding continuous random quantities. In order to understand continuous random quantities, we need a little calculus.

2.1 The probability density function

If X is a continuous random variable, then there exists a function $f_X(x)$, called the *probability density function* (PDF), which satisfies the following:

1. $f_X(x) \geq 0$, for any x,
2. $\int\limits_{-\infty}^{\infty} f_X(x)dx = 1$,
3. $P(a \leq X \leq b) = \int\limits_{-\infty}^{\infty} f_X(x)dx$ for any a and b.

Example. The manufacturer of a certain kind of light bulb claims that the lifetime of the bulb in hours, X can be modeled as a random quantity with PDF

$$f_X(x) = \begin{cases} 0, & x < 100 \\ \frac{c}{x^2}, & x \geq 100, \end{cases}$$

where c is a constant. What value must c take in order for this to define a valid PDF? What is the probability that the bulb lasts no longer than 150 hours? Given that a bulb lasts longer than 150 hours, what is the probability that it lasts longer than 200 hours?

Note

1. Remember that PDFs are *not* probabilities. For example, the density can take values greater than 1 in some regions as long as it still integrates to 1.

2. It is sometimes helpful to think of a PDF as the limit of a relative frequency histogram for many realizations of the random quantity, where the number of realizations is very large and the bin widths are very small.

3. Because $P(X = a) = 0$, we have $P(X \leq k) = P(X < k)$ for continuous random variables.

2.2 The distribution function

In the section 1.2, we defined the cumulative *distribution function* of a random variable X to be

$$F_X(x) = P(X \leq x),$$

for any x. This definition works just as well for continuous random quantities, and is one of the many reasons why the distribution function is so useful. For a discrete random quantity we had

$$F_X(x) = P(X \leq x) = \sum_{\{y \in S_X | y \leq x\}} P(X = y).$$

but for a continuous random quantity we have the continuous analogue

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(-\infty \leq X \leq \infty) \\ &= \int_{-\infty}^{x} f_X(z) dz. \end{aligned}$$

Just as in the discrete case, the distribution function is defined for all $x \in \mathbb{R}$, even if the sample space SX is not the whole of the real line.

Properties

1. Since it represents a probability, $F_X(x) \in [0, 1]$.

2. $F_X(-\infty) = 0$and $F_X(\infty) = 1$.

3. If $a < b$, then $F_X(a) \leq F_X(b)$. ie. F_X is a non-decreasing function.
4. When X is continuous, $F_X(x)$ is *continuous*. Also, by the Fundamental Theorem of Calculus, we have
$$\frac{d}{dx} F_X(x) = f_X(x),$$
and so the slope of the CDF $F_X(x)$ is the PDF $f_X(x)$.

Example. For the light bulb lifetime, X, the distribution function is

$$F_X(x) = \begin{cases} 0, & x < 100 \\ 1 - \frac{100}{x}, & x \geq 100. \end{cases}$$

2.3 Expectation and variance of continuous random quantities

The expectation or mean of a continuous random quantity X is given by

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx,$$

which is just the continuous analogue of the corresponding formula for discrete random variables. Similarly, the *variance* is given by

$$\begin{aligned} V(X) &= \int_{-\infty}^{\infty} [x - E(x)]^2 f_X(x) dx \\ &= \int_{-\infty}^{\infty} x^2 f_X(x) dx - [E(X)]^2 . \end{aligned}$$

So the variance is just

$$V(X) = E[X - E(X)]^2 = E\left(X^2\right) - [E(X)]^2$$

as in the discrete case. Note also that all of the properties of expectation and variance derived for discrete random quantities also hold true in the continuous case.

2.4 The uniform distribution

Now that we understand the basic properties of continuous random variables, we can look at some of the important standard continuous probability models. The simplest of these is the uniform distribution. The random variable X has a uniform distribution over the range $[a, b]$, written

$$X \sim U(a, b)$$

if the PDF is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & otherwise. \end{cases}$$

The expectation of a uniform random quantity is

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \int_{-\infty}^{a} x f_X(x) dx + \int_{a}^{b} x f_X(x) dx + \int_{b}^{\infty} x f_X(x) dx \\ &= 0 + \int_{a}^{b} \frac{x}{b-a} dx + 0 \\ &= \frac{a+b}{2}. \end{aligned}$$

For the variance of X, first we calculate$E\left(X^2\right)$ as follows:

$$E\left(X^2\right) = \int_{a}^{b} \frac{x^2}{b-a} dx = \frac{b^2 + ab + a^2}{3}.$$

Therefor

$$V(X) = E(X^2) - E(X)^2 = \frac{(b-a)^2}{12}.$$

The uniform distribution is rather too simple to realistically model actual experimental data, but is very useful for computer simulation, as random quantities from many different distributions can be obtained from $U(0, 1)$ random quantities.

2.5 The exponential distribution

The random variable X has an exponential distribution with parameter $\lambda > 0$, written

$$X \sim Exp(\lambda)$$

if it has PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & otherwise. \end{cases}$$

The cumulative distribution function,$F_X(x)$ is therefore given by

$$F_X(x) = P(X \leq x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

The expectation of the exponential distribution is

$$E(X) = \int_0^\infty x \lambda e^{-\lambda x} dx = \frac{1}{\lambda},$$

and the variance

$$V(X) = E(X^2) - E(X)^2 = \frac{1}{\lambda^2}.$$

Note

1. As λ increases, the probability of small values of X increases and the mean decreases.
2. The exponential distribution is often used to model lifetime and times between random events. The reasons are given below.

Relationship with the Poisson process

The exponential distribution with parameter λ is the time between events of a Poisson process with rate λ. Let X be the number of events in the interval $(0, t)$. We have seen previously that $X \sim Poi(\lambda t)$. Let T be the time to the first event. Then cumulative distribution function, CDF,

$$\begin{aligned} F_T(t) &= P(T \leq t) \\ &= 1 - P(T > t) \\ &= 1 - P(X = 0) \\ &= 1 - \frac{(\lambda t)^0 e^{-\lambda t}}{0!} \\ &= 1 - e^{-\lambda t}. \end{aligned}$$

This is the distribution function of an $Exp(\lambda)$ random quantity, and so $T \sim Exp(\lambda)$.

Example. Consider again the Poisson process for calls arriving at an ISP at rate 5 per minute. Let T be the time between two consecutive calls. Then we have

$$T \sim Exp(5)$$

and so $E(T) = \sqrt{V(T)} = 1/5$.

2.6 The normal distribution

The normal distribution is the most important and the most widely used among all continuous distribution in the statistics. It is considered as the corner stone of statistics theory. The graph of a Normal distribution is a bell – shaped curved that extends indefinitely in both direction. A random variable X has a normal distribution with parameters μ and σ^2 , written

$$X \sim N(\mu, \sigma^2)$$

if it has probability density function

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad -\infty < x < \infty,$$

for $\sigma > 0$.

Properties of Normal Curve

- The curve is symmetrical about the vertical axis through the mean μ.
- The mode is the highest point on the horizontal axis where the curve is maximum and occurs where $x = \mu$.
- The normal curve approaches the horizontal axis asymptotically.
- The total area under the curve is one (1) or 100%.
- About 68% of all the possible x- values (observations) lie between $\mu - \sigma$ and $\mu + \sigma$, or the area under the curve between $\mu - \sigma$and $\mu + \sigma$ is 68% of the total area.
- About 95% of the observations lie between $\mu - 2\sigma$ and $\mu + 2\sigma$.

- 99.7% (almost all) of the observations lie between $\mu - 3\sigma$ and $\mu + 3\sigma$.
- $E(X) = \mu$ and $V(X) = \sigma^2$.

2.7 The standard normal distribution

A standard normal random quantity is a normal random quantity with zero mean and variance equal to one. It is usually denoted Z, so that

$$Z \sim N(0, 1).$$

Therefore, the density of Z, which is usually denoted $\varphi(z)$, is given by

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \qquad -\infty < z < \infty.$$

It is important to note that the PDF of the standard normal is symmetric about zero. The standard normal distribution is important because it is easy to transform any normal random quantity to a standard normal random quantity by means of a simple linear scaling. Consider $Z \sim N(0, 1)$ and put $X = \mu + \sigma Z$, for $\sigma > 0$. Then $X \sim N(\mu, \sigma^2)$.

Example. Suppose the current measurement in a strip of wire assumed to be normally distributed with a mean of 10mA and a variance of 4mA. What is the probability that the current is greater than 13?

Solution. Let X denote the current in milliamp, the required probability is $P(X > 13)$. Let $Z = \frac{x-\mu}{\sigma} = \frac{13-10}{2} = 1.5$. Therefore $P(X > 13) = P(Z > 1.5) = 0.0668mA$. Here the value obtained from the standard normal table.

Note

The normal distribution is probably *the* most important probability distribution in statistics. In practice, many measured variables may be assumed to be approximately normal. For example, weights, heights, IQ scores, blood pressure measurements etc. are all usually assumed to follow a normal distribution. Common application of the normal distribution is due to the *Central Limit Theorem*. Essentially, this says that *sample means* and *sums* of independent random quantities are approximately normally distributed whatever the distribution of the original quantities, as long as the sample size is reasonably large — more on this in “Applied Statistics”- MA 3102.

Normal approximation of the binomial

If $X \sim B(n, p)$ then X will be well approximated by a Normal distribution if n is large, and p is not too extreme (if p is very small or very large, a Poisson approximation will be more appropriate). A useful guide is that if

- $n \geq 30$
- $np > 5$
- $n(1-p) > 5$

then the binomial distribution may be adequately approximated by a normal distribution. And the tools is

$$Z = \frac{X - np}{\sqrt{np(1-p)}}.$$

That is when the n and p of a binomial distribution are appropriate for approximation by a normal distribution, then

$$B(n, p) \simeq N(np, np[1-p]).$$

Example. Reconsider the number of heads X in 100 tosses of an unbiased coin. There $X \sim B(100, 0.5)$, which may be well approximated as

$$X \simeq N(50, 5^2).$$

So, using normal tables we find that $P(40 \leq X \leq 60) \simeq 0.955$ and $P(30 \leq X \leq 70) \simeq 1$. Note that it satisfy $n \geq 30$, $np > 5$ and $n(1-p) > 5$.

Normal approximation of the Poisson

Since the Poisson is derived from the binomial, in some certain circumstances, the Poisson distribution may also be approximated by the normal. It is generally considered appropriate to make the approximation if the mean of the Poisson is bigger than 20. So, if approximation is done by matching mean and variance:

$$X \sim P(\lambda) \simeq N(\lambda, \lambda) \qquad for \quad \lambda > 20.$$

Example. Reconsider the Poisson process for calls arriving at an ISP at rate 5 per minute. Consider the number of calls X, received in 1 hour. We have

$$X \sim P(5 \times 60) = P(300) \simeq N(300, 300).$$

What is the approximate probability that the number of calls is between 280 and 310?