MATH 323: Lecture Notes

Ryan Ordille

March 28, 2012

Starting from after the midterm and break:

1 28 February

1.1 Random variable distributions

Cumulative distribution function (CDF): $F_X(x) = P(X \leq x)$

For discrete random variables: $P(X = x) = P_X(x)$

1.1.1 Binomial distribution

The random variable X has a *binomial distribution* with parameters n and p if

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

for $x = 0, 1, \ldots, n$, where n is a non-negative integer and $0 \leq p \leq 1$.

Note that we should check if:

1. $P_X(x) \geq 0$ (obvious)

2. $\sum_{\text{all x in range}} P_X(x) = 1$ (easy to check)

We write $X \ Bin(n, p)$ (X has the binomial distribution).

How the binomial distribution arises: (the binomial setup)

Firstly, we have a sequence of n independent trials. That is, the outcomes of these trials are mutually independent.

Secondly, each trial can result in exactly one of two possible outcomes: "success" (S) or "failure" (F). We call such trials *Bernoulli trials*.

Thirdly, the probability of success at trial i is constant and equal to p, for every $i = 1, 2, \ldots, n$. For example, in a coin toss, we cannot change the probability of heads halfway through, so the probability of success is constant.

Theorem: Let X be the number of successes observed in these n trials. Then,

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, 2, \ldots, n$$

Proof: (same idea for the genetic mutation example) First, note that the probability of any particular configuration in which there are x successes (and $(n-x)$ failures) is just $p^x (1-p)^{n-x}$.

E.g:

$$P(S_1 \cap S_2 \cap \ldots \cap S_x \cap F_{x+1} \cap F_{x+2} \cap \ldots \cap F_n) = P(S_1)P(S_2)\ldots P(S_x)P(F_{x+1})\ldots P(F_n)$$

Assuming these are all independent. This is equal to

$$pp\ldots p(1-p)\ldots(1-p) = p^x(1-p)^{n-x}$$

But,

$$\{X = x\} = \cup_{\text{all possible configurations}}\{\text{configuration } i \text{ with } x \text{ successes}\}$$

Note that the configurations are disjoint, so therefore, by axiom 3:

$$P(X = x) = \sum_{\text{all configurations}} P(\text{Configuration}i) = \sum p^x(1-p)^{n-x}$$

Since each configuration is defined by a choice of x objects from n total objects, then for every $x = 0, 1, \ldots, n$:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

2

Example: Suppose that the five year survival probability for lung cancer is .10. If thirty people with lung cancer are sampled, what is the probability that at least three will survive five years or longer?

Solution: Let Y = the number out of thirty who will survive five or more years. We shall reasonably assume the binomial setup, with $n = 30$ and $P(S_i) = .10$, where S_i is the event where the ith subject survives five or more years. Therefore,

$$P(Y \geq 3) = \sum_{y=3}^{30} \binom{30}{y}(.10)^y(1 - .10)^{30-y} = 1 - \sum_{y=0}^{2} \binom{30}{y}(.10)^y(1 - .10)^{30-y}$$

Important note: do not immediately have the "burning desire" to use the binomial distribution as soon as you see a bunch of trials, each of which can result in exactly one of two outcomes. You must check if the trials are independent – they will not be if you are sampling without replacement!

Observe that the Bernoulli distribution is a special case of the binomial distribution, where $n = 1$. So, $P_X(x) = p^x(1 - p)^{1-x}$ for $x = 0, 1$.

1.1.2 Poisson distribution

The random variable X is said to have a *Poisson distribution* with parameter $\lambda > 0$ if

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{for } x = 0, 1, 2, \ldots$$

(Note the infinite range of x)

Check that:

1. $P_X(x) \geq 0$ – obvious

2. $\sum_0^\infty P_X(x) = 1$ – Taylor series expansion for e^λ

The Poisson distribution arises as an approximation to the binomial for "large n and small p".

Theorem: let X $Bin(n, p)$, then the limit of $P(X = x)$ as n tends to infinity and p tends to 0, in such a way that $n \times p$ is constant $(= \lambda)$, is

$$\frac{\lambda^x e^{-\lambda}}{x!} \quad \text{for } x = 0, 1, 2, \ldots$$

(e.g. $p = \frac{6}{n} : np = n(\frac{6}{n}) = 6$)

Proof: We shall use the following result in proving our theorem.

$$\lim n \to \infty (1 + \frac{a}{n})^n = e^a$$

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \text{ for } x = 0, 1, \ldots, n$$

$$(*) = \frac{n!}{x!(n - x)!} p^x (1 - p)^n (1 - p)^{-x}$$

Now since $\lambda = np$, we have $p = \frac{\lambda}{n}$. Then, $(*)$ becomes:

$$\frac{n!}{x!(n - x)!} \frac{\lambda^x}{n^x} (1 - \frac{\lambda}{n})^n (1 - \frac{\lambda}{n})^{-x}$$

After some algebraic manipulation, this becomes:

$$\frac{1}{x!} \frac{n(n - 1)(n - 2) \ldots (n - x + 1)}{nn \ldots n} (1 - \frac{\lambda}{n})^n (1 - \frac{\lambda}{n})^{-x} \lambda^x$$

Notice that the second term has x terms on both the top and the bottom. Now, let n go to infinity to get the limit:

$$\frac{\lambda^x}{x!} e^{-\lambda} \blacksquare$$

2 01 March

2.1 Poisson distribution continued

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

Example: Suppose that in a book of 1000 pages, on any particular page, there can be either zero errors or one error. Suppose further that the probability of an error on any particular page is .002. what is the approximate probability that there will be at most three errors in the book?

Exact solution: There is a binomial setup – errors are likely to occur independently amongst the $n = 1000$ trials. Each trial can result in either a success (an error is

found) or a failure (an error is not found). There is also a constant probability of success (.002) for every trial. Let X be the number of successes in 1000 pages. Then $X \sim Bin(1000, .002)$.

$$P(X \leq 3) = \sum_{x=0}^{3} \binom{1000}{x} (.002)^x (1 - .002)^{1000-x}$$

Approximate solution: Since n is large and p is small, we can use the Poisson approximation with $x = n \times p = 1000 \times \frac{2}{1000} = 2$. Therefore:

$$P(X \leq 3) = \sum_{x=0}^{3} P(X = x) = \sum_{x=0}^{3} \frac{2^x e^{-2}}{x!}$$

If X has a Poisson distribution with parameter λ, we write $X \sim Po(\lambda)$.

2.2   The Hypergeometric Distribution

The random variable X has a hypergeometric distribution with parameters N, a, and n if:
$$P(X = x) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{Nn}$$
(for $x \leq a$ and $n - x \leq N - a$)

2.3   The Geometric Distribution

X is said to have a geometric distribution with parameter p if $P(X = x) = P_x(x) = (1-p)^{x-1}p$ for $x = 1, 2, \ldots$.

$$p \sum_{x=1}^{\infty} (1-p)^{x-1} = p \frac{1}{1 - (p-1)} = 1$$

The geometric r.v. is used to describe or model the trial number at which the first success occurs in a sequence of independent Bernoulli trials, each with the probability of success p.

2.4 Expected values

The probability distribution of a random variable provides the complete story about the r.v.. There is no information about a r.v. once we know its probability distribution. However, we often wish to summarize a probability distribution. The two most common summaries are:

1. a parameter that discusses the "centre of distribution" and

2. a parameter that discusses how spread out the values are from the centre.

To this end, we give the following definition:

Expected value: let Y be a discrete r.v. with the probability function $P_Y(y)$. Then, we define the expected value of y denoted by $E(Y)$ to be

$$E(Y) = \sum_{\text{all } y} y P_Y(y) = \sum_{\text{all } y} y P(Y = y)$$

3 06 March

3.1 Review

Recall: given a r.v. Y, we define the expected value or expectation of Y denoted by $E(Y)$ to be:

$$E(Y) = \sum_{\text{all } y} y P(Y = y)$$

provided that the sum is finite.

3.2 Theory

Notes:

1. $E(Y)$ is often denoted by μ_y, and also is termed the *mean* of Y (also called the "population mean").

2. **Interpretation:** $E(Y)$ is a weighted average or mean of the possible values of the random variable Y, where the weights are the probabilities of these values. For example, in the special case where Y has a discrete uniform distribution

$a_1, a_2, \ldots, a_N$, then

$$E(Y) = \sum_{i=1}^{N} a_i P(Y = a_i) = \frac{1}{N} \sum_{i=1}^{N} a_i$$

So think of μ as the "average value" of Y.

3. μ is a constant, a parameter specific to a given probability distribution. You need the distribution to compute μ.

4. $E(cY) = cE(Y)$ where c is a constant, and $E(\sum_{i=1}^{n} Y_i) = \sum_{i=1}^{n} E(Y_i)$ (the proof will come later)

5. Note, however, that in general, $E(XY) \neq E(X)E(Y)$.

6. Let g be some real-valued function of a r.v. Y. Then, if $X = g(y)$, we have

$$E(X) = E(g(y)) = \sum_{\text{all } y} g(y)P(Y = y)$$

The point of this: by definition, $E(g(y)) = E(X) = \sum_{\text{all } x} xP(X = x)$. Thus, in order to find $E(g(y))$, it would appear that we first need to find the probability distribution of $X = g(y)$. We'll see such transformations later – they can be difficult. You do not have to first find the distribution of X – you can use the distribution of the original Y and sum $g(y)P(Y = y)$.

In particular, if $g(Y) = Y^k$, we can call $E(g(Y)) = E(Y^k)$ is called the kth moment of Y. μ is called the first moment of Y.

$$E(Y^k) = \sum_{\text{all } y} y^k P(Y = y)$$

Of particular importance is a special function of Y:

$$g(Y) = (Y - \mu_Y)^2$$

In this case, $E(g(Y)) = E((Y - \mu_Y)^2)$ is called the *variance* of Y, and is denoted by $Var(Y)$, also σ_Y^2. This gives the average squared distance between the values of Y and its mean. It is a measure of spread or variation of Y (i.e. its distribution). We call $\sqrt{\sigma_Y^2}$ the *standard deviation* of Y. This is more convenient than $Var(Y)$, since it is in the same units as Y, unlike $Var(Y)$.

Result:
$$Var(Y) = E((Y - \mu_Y)^2) = E(Y^2) - \mu_Y^2$$

Proof:

$$
\begin{aligned}
E((Y - \mu_Y)^2) &= E(Y^2 - 2\mu_Y Y + \mu_Y^2) \\
&= E(Y^2) - E(2\mu_Y Y) + E(\mu_Y^2) \\
&= E(Y^2) - 2\mu_Y E(Y) + \mu_Y^2 \\
&= E(Y^2) - 2\mu_Y^2 + \mu_Y^2 \\
&= E(Y^2) - \mu_Y^2 \ \blacksquare
\end{aligned}
$$

Final note: $Var(cY) \neq c \times Var(Y)$, but $Var(cY) = c^2 \times Var(Y)$.

3.3 Examples

3.3.1 Example 1

The calculation of insurance premiums: An insurance company will insure your computer against theft for $1000. It is known with a probability .05 that your computer will be stolen. What premium should the insurance company charge so that its expected gain is 0?

Solution: Let c be the required premium. Let Y be the gain of the company in a given year. We need to find the value of c such that $E(Y) = 0$.

First, we need $P_Y(y)$ for all y. We have $P(Y = c) = .95$ (where the computer was not stolen) and $P(Y = (c - 1000)) = .05$ (where the computer was stolen). Therefore,
$$E(Y) = c \times .95 + (c - 100) \times .05$$

Setting $E(Y) = 0$ and solving for c, we get $c = 50$. Thus, if the company charges $50 for the policy, on average, over a large number of clients, they would neither lose nor gain money.

3.3.2 Example 2

"Nuts and Bolts" Example: suppose that the r.v. x has the probability distribution:

$$P(X = -1.2) = .32$$
$$P(X = 2.6) = .40$$
$$P(X = 0) = .28$$

Find $E(X)$ and $Var(X)$.

$$E(X) = -1.2 \times .32 + 0 \times .28 + 2.6 \times .40 = .66 = \mu_X$$

$$Var(X) = E(X^2) - \mu_X^2 = \sum x^2 P(X = x)$$
$$E(X^2) = (-1.2)^2 \times .32 + 0^2 \times .28 + (2.6)^2 \times .40$$
$$= 3.1648$$
$$\therefore Var(X) = 3.1648 - (.66)^2$$
$$= 2.7292$$
$$\text{Also, } \sigma = \sqrt{2.7292} = 1.652$$

4   08 March

4.1   Summary

Centre $- E(X) = \mu$

Spread $- Var(x) = \sigma^2$

Standard Deviation $- \sqrt{\sigma^2} = \sigma$

$$\sum (x - \mu)^2 P(X = x) = \sum x^2 P(X = x) - \mu^2$$

4.2 The Mean and Variance of Some Named Distributions

4.2.1 Binonmial

$$E(X) = \mu = \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum \frac{n(n-1)!}{(x-1)!(n-1-(x-1))!} p p^{x-1} (1-p)^{n-1-(x-1)} \quad \text{as } (n-x)! = (n-1-(x-1))!$$

$$= np \sum \binom{n-1}{x-1} p^{x-1} (1-p)^{n-1-(x-1)}$$

$$= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y}$$

$$= np$$

The sum above is equal to 1 since it is just a sum of $n-1$ $Bin(n-1, p)$ probabilities.

To find $Var(X)$, we first have to find $E(X^2)$. By definition:

$$E(X^2) = \sum_{x=1}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x}$$

Note that x^2 will not cancel with the leading terms of $x!$ as before, so we have to use a trick. We first calculate $E(X(X-1))$, which is easy. Then, notice that

$$E(X(X-1)) = E(X^2) - E(X) = E(X^2) - \mu$$

So, $E(X^2) = E(X(X-1)) + \mu$. Finally, $Var(X) = E(X(X-1)) + \mu - \mu^2$.

$$E(X(X-1)) = \sum_{x=2}^{n} x(x-1) \binom{n}{x} p^x (1-p)^{n-x}$$

$$= n(n-1)p^2 \sum \frac{(n-2)!}{(x-2)!(n-2-(x-2))!} p^{x-2} (1-p)^{n-2-(x-2)}$$

$$= n(n-1)P^2 \sum \binom{n-2}{x-2} p^{x-2} (1-p)^{n-2-(x-2)}$$

$$= n(n-1)p^2$$

$$Var(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p)$$

4.2.2 Bernoulli

In particular, the mean and variance of a Bernoulli r.v. are p and $p(1-p)$ respectively (because $n=1$).

4.2.3 Poisson

$$E(X) = \sum_{x=1}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!}$$
$$= e^{-\lambda} \lambda \sum_{x-1=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$
$$= \lambda$$

So, the mean of a Poisson r.v. is just its parameter λ.

For the variance, first find $E(X(X-1))$. We find $Var(X) = \lambda$, the same as the mean.

4.2.4   Geometric

$$E(X) = \sum_{x=1}^{\infty} xp(1-p)^{x-1}$$

using our trick...
$$= p\sum_{x=1}^{\infty} x(1-p)^{x-1}$$

$$= p\sum_{x=1}^{\infty} -\frac{d}{dp}(1-p)^x$$

$$= -p\frac{d}{dp}\sum_{x=1}^{\infty}(1-p)^x \text{ we can interchange the derivative and the sum}$$

$$= -p\frac{d}{dp}\frac{1-p}{1-(1-p)} \quad \text{where } (1-p) = r$$

$$= -p\frac{d}{dp}(\frac{1}{p}-1)$$

$$= \frac{1}{p}$$

For the variance, first find $E(X(X-1))$ (2 derivatives), and then you can find the variance.
$$Var(X) = \frac{1-p}{p^2}$$

4.3   Continuous probability distributions

Defintion: a r.v. X with cdf F_X is said to be continuous if F_X is continuous for all $-\infty < x < \infty$.

The continuous cdfs split into two types:

1. the so-called "absolutely continuous" cdfs (essentially, they are differentiable) and

2. the so-called "singular" cdfs (without derivatives).

From now on in this course, we'll assume for continuos cdfs F_X, that they are differentiable for all $-\infty < x < \infty$.

12

It follows that if X is continuous, then $P(X = x) = 0$ for every x. Consider the "area under a curve" analogy.

Remember (as $F_X(x) = P(X \leq x)$):

$$P(X = x) = F_X(x) - P(X < x)$$

Hence, we cannot specify a continuous r.v. by specifying the values $P(X = x)$ for all x that X can assume, as we did in the discrete case. Instead, we introduce an analogue of the probability function known as the *probability density function* (pdf). It will turn out that the pdf also uniquely determines the probability distribution.

Definition: A real-valued function f_x is said to be the probability density function of a r.v. X if:

1. $f_x(x) \geq 0$ for all $-\infty < x < \infty$ and

2. $P(X \in A) = \int_A f_x(x)dx$ (i.e. f_x has the property that when you integrate it over a set, you get the probability of the set).

5   13 March

5.1   The probability density function

Definition: the probability density function (p.d.f.) of a r.v. X is any function $f_X(x)$ such that:

1. $f_X(x) \geq 0 \ \forall -\infty < x < \infty$ and

2. $P(X \in A) = \int_A f_X(s)dx \ \forall$ events $A \in [R]$

A pdf gives $P(X \in A)$ by integrating over A.

5.2   Notes on the p.d.f.

(1) In particular, if A is of the form $(-\infty, x]$ then

$$P(X \in A) = P(X \leq x) = \int_{-\infty}^{x} f_X(y)dy$$

13

In short, $F_X(x) = \int_{-\infty}^{x} f_x(y)dy$.

(2) Conversely, we can recover the p.d.f. from the c.d.f. by the Fundamental Theorem of Calculus, since:

$$\frac{d}{dx}F_X(x) = F_X(x) = f_X(x) \forall x$$

Thus, in particular, if f_X is a p.d.f., then:

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

(3) Interpretation of a p.d.f: We have, for small Δx, that:

$$\frac{F_X(x + \delta x) - F_X(x)}{\Delta x} \approx f_X(x)$$

So it follows that:

$$F_X(x + \delta x) - F_X(x) \approx \Delta x f_X(x)$$

But the left hand side is just $P(x < X \leq x + \Delta x)$. Finally, we have that $f_X(x)\Delta x$ is approximately the probability that X lies in $(x, x + \Delta x]$.

Note that, because the p.d.f. does not represent a probability on its own, it can be greater than 1 or less than 0. When multiplied by a small Δx, it gives an *approximate* probability.

Any function whose total area equals 1 can qualify as a p.d.f., even if some points are larger than 1.

(4) For continuous r.v.s, we define:

$$E(g(X)) = \sum_{\text{all}x} g(x)P_X(x) = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

As before, when $g(X) = X^k$, then we refer to the $E(g(X)) = E(X^k)$ as the kth moment. Of particular importance are the first moment $(k = 1)$ and the second moment $(k = 2)$. Again, as before, we call the first moment the *mean* or the *expected value* of X.

14

So, by definition:
$$E(X) = \mu_X = \int_{-\infty}^{\infty} x f_X(x) dx$$

and
$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_x(x) dx)$$

Finally, as before:
$$Var(X) = E((X - \mu_X)^2) = \int_{\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx = E(X^2) - \mu_X^2$$

and
$$Var(X) = \int_{-\infty}^{\infty} x^2 f_X(x) dx - (\int_{-\infty}^{\infty} x f_X(x) dx)^2$$

Note: watch out for a p.d.f. that may change its form in different ranges of $(-\infty, \infty)$ when carrying out an integration.

5.3   Examples

5.3.1   Example 1

Let $f_X(x) = c(x^2 + 1)$ for $0 < x < 1$ and $f_X(x) = 0$ elsewhere (c is a constant).

1. Find c.

2. Find $P(.25 < X \le .50)$.

3. Find $P(.25 < X < .50)$.

4. Find F_X.

5. Find $E(X)$ and σ_X.

(1) Since $\int_{-\infty}^{\infty} f_X(x) dx = 1$, we must have:

$$\int_{-\infty}^{0} 0 dx + \int_{0}^{1} c(x^2 + 1) dx + \int_{1}^{\infty} 0 dx = 1$$

This gives $c = .75$.

15

(2)

$$P(.25 < X \le .50) = \int_{.25}^{.50} (.75)(x^2 + 1)dx = \frac{55}{256}$$

(3) For a continuous p.d.f., these two values are the same (by rules of integration).

$$P(.25 < X < .50) = P(.25 < X \le .50) = \frac{55}{256}$$

(4)

$$F_X(x) = 0 \ \forall x \le 0$$

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy \ \forall 0 < x < 1$$

$$= \int_{-\infty}^{0} 0 dy + \int_{0}^{x} (.75)(y^2 + 1)dy$$

$$= (.75)(\frac{x^3}{3} + x)$$

$$F_X(x) = 1 \ \forall x \ge 1$$

(5)

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

$$= \int_{0}^{1} x(.75)(x^2 + 1)dx$$

$$= (.75)(\frac{x^4}{4} + \frac{x^2}{2})(0 \text{ to } 1)$$

$$= (.75)(.25 + .50)$$

$$= \frac{9}{16}$$

To find σ_X, first find:

$$E(X^2) = \int_{-\infty}^{0} 0dx + \int_{0}^{1} x^2(.75)(x^2+1)dx + \int_{1}^{\infty} 0dx$$
$$= (.75)(\frac{x^5}{5} + \frac{x^3}{3} \text{ 0 to 1}$$
$$\sigma^2 = Var(X)$$
$$= E(X^2) - (\frac{9}{16})^2$$
$$\sigma = \sqrt{\sigma^2}$$

6   15 March

6.1   Named continuous probability distributions

6.1.1   The uniform distribution

Definition: The r.v. X is said to be uniformly distributed on the interval $[a, b]$ if

$$f_X(x) = \int_{a}^{x} \frac{1}{b-a} \quad \text{for } a \le x \le b$$

and $f_X(x) = 0$ elsewhere.

Notes:

1. The p.d.f. is constant on $[a, b]$. On the graph, the height is constantly $\frac{1}{b-a}$, so the area under the curve is 1. The probability is uniformly spread out in the interval.

2. The uniform distribution is often used to model situations in which we believe outcomes occur completely at random.

3. An important special case is the uniform distribution $[0, 1]$.

4. Notation – we write $X \, U[a, b]$ to mean that X has a uniform distribution on the interval $[a, b]$.

5. The c.d.f. of X is:

$$F_X(x) = 0 \text{ for} x < a$$
$$= \frac{x-a}{b-a} \text{ for} a \le x \le b$$
$$= 1 \text{ for} b < x$$

The graph of the c.d.f. is 0 up to a, then grows linearly up to b, then is constantly 1 after.

6.

$$\mu = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$
$$= 0 + \int_a^b x \frac{1}{b-a} dx + 0$$
$$= \frac{b^2 - a^2}{2(b-a)}$$
$$= \frac{a+b}{2}$$

It is easy to see that the variance of a uniform distribution is $Var(X) = \frac{(b-a)^2}{12}$.

6.1.2   The exponential distribution

Definition: X has an exponential distribution with parameter $\beta > 0$ if:

$$f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} \text{ for } x \ge 0$$

The distribution is equal to 0 elsewhere.

Notes:

1. we write $X \sim Exp(\beta)$ to describe this distribution.

2. If X is exponential, then X is a non-negative r.v., meaning $P(X \ge 0) = 1$.

3. The p.d.f. of $f_X(x)$ is $\frac{1}{\beta}$ at $x = 0$, and grows exponentially downwards as x grows larger. The probability of an interval of length L decreases as L moves down the graph (e.g. the probability between 2 and 4 is greater than the probability between 4 and 6). The probability is concentrated towards the origin.

18

4. The c.d.f. of X is:

$$F_X(x) = 0 \quad \forall x < 0$$

$$= 0 + \int_0^x \frac{a}{\beta} e^{\frac{-y}{\beta}} dy \quad \text{for } x \geq 0$$

$$= 1 - e^{\frac{-x}{\beta}}$$

5. $\mu = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = 0 + \int_0^\infty x \frac{1}{\beta} e^{-\frac{x}{\beta}} dx = \beta$ You can do this using integration by parts, or we'll see a trick for this later. Also, $\sigma^2 = Var(X) = \beta^2$.

6. Sometimes, the exponential distribution will be "parameterized" in a different way, i.e. the parameter will be written in a different form. The alternative form for the p.d.f. is:

$$f_X(x) = \lambda e^{\lambda x} \text{ for} x \geq 0$$

$$= 0 \text{ elsewhere}$$

Watch out how the writer is writing in the parameter for the distribution! In this case, $E(X) = \frac{1}{\lambda}$ and $Var(X) = \frac{1}{\lambda^2}$.

7. **The memoryless property:**
Theorem: Let X $Exp(\beta)$. Then, $P(x \leq X < x + h | X \geq x) = P(0 \leq X < h)$. In other words, the information that $X \geq x$ is "forgotten".
Important note: the memoryless property *does not* assert that $P(0 \leq X < h) = P(x \leq X < x + h)$!
Proof: Let $x \leq X < x + h$ be B, and $X \geq x$ be A. Then,

$$P(x \leq X < x + h | X \geq x) = P(B \cap A)/P(A)$$

$$= P(B)/P(B) \text{since B is a subset of A}$$

$$= \frac{F_X(x + h) - F_X(x)}{1 - P(X < x)}$$

$$= \frac{1 - e^{\frac{-(x+h)}{\beta}} - (1 - e^{\frac{-x}{\beta}})}{1 - (1 - e^{\frac{-x}{\beta}})}$$

$$= 1 - e^{\frac{-h}{\beta}}$$

$$= P(0 < X \leq h)$$

There is an interesting converse – the *only* continuous distribution with the memoryless property is the exponential. The geometric discrete distribution also has this property.

19

8. The exponential distribution is used when you believe that X has a constant "hazard" (i.e. $P(x \leq X < x + h | X \geq x)$ is roughly constant in x for small h). It is also used to model the times between events that occur according to a Poisson Process (explained in later statistic courses).

7 20 March

7.1 The Gamma Distribution

Before defining the gamma distribution, we need to define the *gamma function*.

7.1.1 The Gamma Function

Let $\alpha > 0$. We denote the gamma function by $\Gamma(\alpha)$ and define

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \, dx$$

The gamma function has the following two important properties:

1. $\Gamma(\frac{1}{2}) = \sqrt{(\pi)}$
2. $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$

7.1.2 The Gamma Distribution

Definition: A r.v. X has a gamma density with parameters α, β if its p.d.f. is given by:

$$f_X(x) = \frac{1}{\Gamma(\alpha)} \frac{x^{\alpha-1} e^{\frac{-x}{\beta}}}{\beta^\alpha} \text{for } x \geq 0$$

$f_X(x) = 0$ when $x < 0$.

Notes:

(1)

$$\int_0^\infty \frac{1}{\Gamma(\alpha)} \frac{x^{\alpha-1} e^{\frac{-x}{\beta}}}{\beta^\alpha} \, dx = 1$$

This is as it should be – set $y = \frac{x}{\beta}$ to get $\frac{1}{\Gamma(\alpha)} \int_0^\infty y^{\alpha-1} e^{-y} \, dy = 1$. Set $y = \frac{x}{\beta}$ and $dx = \beta \, dy$ to get the same answer.

This just proves that this given formula is a density, as it integrates out to 1.

(2) The gamma distribution is said to "flexible", meaning that many different shapes for the p.d.f. can be induced by changing the two parameters α and β.

For $\alpha > 1$, the p.d.f. grows rapidly immediately after $x > 0$, then drops off with a tail as x grows larger. $f_X(\alpha > 1)$ is said to be skewed to the right.

For $\alpha = 1$, $f_X(\alpha = 1$ grows exponentially downwards, similar to the exponential distribution.

For $\alpha < 1$, the p.d.f. also grows exponentially downwards, but more steeply.

(3) The gamma density can be used to model the waiting time for the nth event if the times between events are independent exponential r.v.s.

(4) There are two important special cases of the gamma distribution:

1. If we set $\alpha = 1$, we get an exponential distribution with parameter β.

2. If we set $\alpha = \frac{\nu}{2}$ and $\beta = 2$, then the density becomes:

$$\frac{1}{\Gamma(\frac{\nu}{2})} \frac{x^{\frac{\nu}{2}-1}}{2^{\frac{\nu}{2}}} e^{\frac{-x}{2}}$$

For $x \geq 0$ (0 otherwise).
This particular p.d.f. plays an important role in statistics, and is called a *Chi-square* p.d.f. "with ν degrees of freedom". ν is just a parameter with this peculiar name.
We write $X \sim \chi_\nu^2$ to mean "X has a Chi-square distribution with ν degrees of freedom".

(5) It is not too difficult to derive $E(X)$ and $Var(X)$ from the definition. It will be easier, however, once we know about moment-generating functions.

In the end, $E(X) = \alpha\beta$. Write x^α as $x^{\alpha+1-1}$ and let $y = \frac{x}{\beta}$, and carry out the integration.

We get, similarly, $Var(X) = \alpha\beta^2$.

In particular, if $X \sim \chi_\nu^2$, then $E(X) = \nu$ (set $\alpha = \frac{\nu}{2}$ and $\beta = 2$) and $Var(X) = 2\nu$).

(**6**) The c.d.f. F_X is not known in closed form: $F(x) = 0$ for $x < 0$ and:

$$F_X(x) = \int_0^x \frac{1}{\Gamma(\alpha)} \frac{y^{\alpha-1}}{\beta^\alpha} e^{\frac{-y}{\beta}} \, dy$$

For $x > 0$.

(**7**) Notation:

We write $X \; Gamma(\alpha, \beta)$ to mean "X has a gamma distribution with parameters α, β".

7.2 The Normal (or Gaussian) Distribution

The normal or Gaussian distribution is easily the most important distribution in probability and statistics! The distribution seems to occur naturally all over.

7.2.1 Definition

The r.v. X has a normal distribution with parameters μ, σ^2 if its p.d.f. is given by

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

for $-\infty < x < \infty$.

7.2.2 Notes

(**1**) We write $X \sim N(\mu, \sigma^2)$.

(**2**) It is possible to show directly that $E(X) = $ the parameter μ and $Var(X) = $ the parameter σ^2.

Note that if $X \sim N(1.2, 7.8)$, we mean that $\mu = 1.2$ and $\sigma^2 = 7.8$ *not* $\sigma = 7.8$. We'll derive μ and σ^2 by using the so-called moment generating function later, as the current integration would be a bit tricky (but not impossible).

(**3**) The p.d.f. has the famous bell shape. The features are:

1. f_X is symmetric about μ.

2. Changing μ changes the location of the p.d.f., i.e. where it is centred on the x-axis.

3. Increasing σ^2 increases the spread of the p.d.f. and decreasing σ^2 decreases the spread.

(4) The c.d.f. is not known in closed form, similar to the gamma distribution. Probabilities of intervals need to be done using numerical integration.

However, unlike the gamma density, it is possible to use a single table to find any normal probability. The idea is to reduce the general problem to what is called a *standard normal problem*.

7.2.3 Standard Normal Problem

Background: if I give you *any* r.v. with mean μ and standard deviation σ, then
$$Y = \frac{X - \mu}{\sigma}$$
has $E(Y) = 0$ and $Var(Y) = 1$.

We are said to have *standardized X*.

However, if $X \sim N(\mu, \sigma^2)$, we have the following:
$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$
This is called a *standard normal* r.v. or distribution.

8   22 March

8.1   Standardizing continued

Our main result from Tuesday: if $X \sim N(\mu, \sigma^2)$, then
$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Note: For any r.v. with mean μ and variance σ^2, $\frac{X-\mu}{\sigma}$ will have mean 0 and variance 1. The proof of this is simple - just plug in the fraction for X in $E(X)$ and $Var(X)$.

The point of the main result is that, after standardizing, we still get a normal r.v. If you standardize a r.v., you don't always get a r.v. of the same type (unless you're standardizing a r.v. with a normal distribution).

8.1.1 Example 1

Problem: If $X \sim N(-1.2, 4)$, find $P(-1.9 \leq X < 2.2)$.

Solution: The idea is to reduce the problem to a $N(0, 1)$ problem, and then use $N(0, 1)$ tables.

Step 1: (do not draw a sketch now)

$$P(-1.9 \leq X < 2.2) = P(\frac{-1.9 - (-1.2)}{2} \leq \frac{X - (-1.2)}{2} < \frac{2.2 - (-1.2)}{2})$$
$$= P(-.35 \leq Z < 1.7) \quad \text{our main result from before}$$

Step 2: draw a sketch: (draw a standard normal distribution with mean $= 0$, shade in area A between $-.35$ and 1.7)

Tables will give you areas to the right of a value z. Areas to the right of $z = 3$ is essentially 0, so tables will usually not give values of $z > 3$. Recall that the areas will be the probabilities – e.g. the area to the right of $z = 1.78$ will equal $P(Z \geq 1.78)$.

We'll call A_1 the area between $-.35$ and 0, and A_2 the area between 0 and 1.7. We'll get these values by using the symmetry of $N(0, 1)$ about $\mu = 0$. Tables only give positive values of z, so to get A_1, subtract $P(Z \geq .35)$ from .5.

From the table values, $A_2 = .5 - .0446$ and $A_1 = .5 - .3632$, so $A = A_1 + A_2 = .5922$.

8.1.2 Example 2

Problem: Use the $N(0, 1)$ tables inversely here. Suppose that a car battery is known to have a lifetime that is approximately normally distributed with a mean of 36 months and a standard deviation of 6 months. What should the warranty period be set at so that only 5% of batteries will need to be replaced?

Notice: Batteries cannot have a negative lifetime, so our true normal distribution will not work. However, our lifetime is so skewed to the right that 2σ is still way to the right of 0, so we can shift our model. Strictly speaking, modelling anything that cannot hold negative values is not correctly, but for almost all cases, the normal distribution will work just fine.

Solution: We have that $X \sim N(36, 6^2)$. Let x_0 be the required warranty period. We want that x_0 such that $P(X < x_0) = .05$, i.e. such that $P(X \geq x_0) = .95$.

Reduce this distribution to a standard normal distribution. So, we seek x_0 such that:
$$P(\frac{X - 36}{6} \geq \frac{x_0 - 36}{6}) = .95$$
i.e. such that $P(Z \geq \frac{x_0 - 46}{6}) = .95$.

Draw a sketch – we're looking for a z_0 from standard normal tables such that the area to the right is .95, then set that equal to our above probability, and we can solve our problem.

This z_0 must be to the left of the mean 0, since the area to the right of the mean is .5. Find a z_1 such that the area to the right of it is .05, and according to our tables, $z_1 = 1.64$. Take the negative, so the area to the right of $z_0 = -1.64$ is .95 (using the symmetry of the normal distribution).

From the $N(0, 1)$ tables, we know that $P(Z \leq -1.64) = .95$. Finally, we must have that
$$\frac{x_0 - 36}{6} = -1.64$$
We get $x_0 = 26.16$ months.

8.2 Moment Generating Functions

8.2.1 Definition

Let X be a r.v. with p.d.f. f_X (respectively, probability P_X for the discrete case). We define the moment generating function (denoted m.g.f.) to be that function of t, such that
$$M_X(t) = E(e^{tx})$$

8.2.2 Notes

(1) The m.g.f. is a function of the real values t.

(2) In the continuous case,

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

In the discrete case,

$$E(e^{tx}) = \sum_x e^{tx} P(X = x)$$

(3) For some distributions, the m.g.f. does not exist because the integral (or sum) does not converge. We say that the m.g.f. exists if it exists in some interval containing 0.

(4) If the m.g.f. exists, then it is possible to recover the p.d.f. or probability function, i.e. there is a one-to-one correspondence between a p.d.f. (pf) and a m.g.f.. Recovering it, although possible, is a bit complicated, and we will not be expected to do so.

(5) Uses of the m.g.f. – the m.g.f. can be used to find the *moments* of a r.v., and is often easier than finding the moments $(E(X^k))$ by using the definition.

Theorem: $E(X^k) = M^{(k)}(0)$.

Proof (continuous case): (discrete case – replace integral by sum) We have, by

26

definition:

$$M_X(t) = E(e^{tx}) = \int_{-\infty} \infty e^{tx} f_X(x) \, dx$$

$$M_X^1(t) = \frac{d}{dt} \int \cdots$$

$$= \int \frac{d}{dt} \cdots$$

$$= \int x e^{tx} f_X(x) \, dx$$

now set $t = 0$

$$M_X^1(t) = \int_{-\infty}^{\infty} x f_X(x) \, dx$$

$$= E(X)$$

In general, we get

$$M^{(k)}(t) = \int_{-\infty} \infty x^k e^{tx} f_X(x) \, dx$$

This gives $M^{(k)}(t) = E(X^l)$, as advertised.

9   27 March

9.1   Recall

$$M_X(t) = E(e^{tx})$$

$$M_X^{(k)}(0) = E(X^k)$$

9.2 The m.g.f.s of some important distributions

9.2.1 Gamma distribution

$X \sim Gamma(\alpha, \beta)$

$$M_X(t) = \int_0^\infty e^{tx} \frac{1}{\Gamma(\alpha)} \frac{x^{\alpha-1}}{\beta^\alpha} e^{\frac{-x}{\beta}} \, dx$$

$$= \frac{1}{\Gamma(\alpha)} \frac{1}{\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x(\frac{1}{\beta}-t)} \, dx$$

Set $y = x(\frac{1}{\beta} - t)$ with $dy = (\frac{1}{\beta}) - t \, dx$ to get

$$M_X(t) = \frac{1}{(1-\beta t)^\alpha} \quad \text{for } |\beta t| < 1$$

From $M_X(t)$, we immediately get

$$M_X'(0) = \alpha\beta(1-\beta t)^{\alpha-1}\Big|_{t=0}$$

by the chain rule. Therefore, $M_X'(0) = \alpha\beta$. Similarly,

$$M_X''(t)\Big|_{t=0} = \alpha\beta^2 + \alpha^2\beta^2$$

as $Var(Y) = E(Y^2) - (E(Y))^2$. Therefore, $Var(X) = \alpha\beta^2$.

So, in particular, for $\alpha = 1$ (i.e. the exponential distribution), we have

$$M_X(t) = \frac{1}{(1-\beta t)}$$

with $E(X) = \beta$ and $Var(X) = \beta^2$.

For $\alpha = \frac{\nu}{2}$ and $\beta = 2$ (i.e. a chi-square distribution with ν degrees of freedom), we get

$$M_X(t) = \frac{1}{(1-2t)^{\frac{\nu}{2}}}$$

and $E(X) = \nu$ and $Var(X) = 2\nu$.

9.2.2 Binomial distribution

$X \sim Bin(n, p)$

$$M_X(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$
$$= (1 - p + pe^t)^n \ \forall t \in (-\infty, \infty)$$

(note: $1 - p = a$ and $pe^t = b$)

We get $M'_X(0) = np$ and $M''_X(0) = np(1-p) + n^2p^2$. Therefore, $Var(X) = np(1 - p)$.

9.2.3 Poisson distribution

$X \sim Po(\lambda)$

$$M_X(t) = E(e^{tx})$$
$$= \sum_{x=0}^{\infty} \frac{e^{tx} \lambda^x e^{-\lambda}}{x!}$$
$$= e^{\lambda(e^t - 1)}$$

Therefore, $E(X) = M'_X(0) = \lambda$ and $E(X^2) = \lambda + \lambda^2$, so therefore $Var(X) = \lambda$.

9.2.4 Normal distribution

Let $X \sim N(0, 1)$ to start with.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{1}{2}x^2} \ \forall x \in (-\infty, \infty)$$

Therefore,

$$
\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \, dx \\
&= \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2 - 2tx + t^2)} e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \, dx \\
&= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-t)^2} \, dx \\
&= e^{\frac{t^2}{2}} \text{ since the integrand is just a } N(t,1) \text{ p.d.f.}.
\end{aligned}
$$

To get the m.g.f. of a $N(\mu, \sigma^2)$ r.v. for arbitrary μ and σ^2, we use the following property of an m.g.f.: let a and b be constants. Then,

$$
M_{aX+b}(t) = e^{bt} M_X(at)
$$

Now, recall that if $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$. Therefore, we can always write a $N(\mu, \sigma^2)$ r.v. X as $X = \sigma Z + \mu$.

Putting these two results together, we get

$$
M_X(t) = M_{\sigma Z + \mu}(t) = e^{\mu t} e^{\frac{\sigma^2 t^2}{2}}
$$

and

$$
M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \quad \forall t \in (-\infty, \infty)
$$

We find $M_X'(0) = \mu$, $M_X''(0) = \sigma^2 + \mu^2$, and $Var(X) = \sigma^2$.

9.3   Transformations of random variables

Often, we're given the distribution of a r.v. X, but we're more interested in some function $Y = g(X)$ of this r.v., e.g. maybe we have the distribution of the velocity v, and we are interested in the distribution of the kinetic energy $Y = \frac{1}{2}mv^2$. In general, we're concerned with finding the distribution of $g(X)$ knowing the distribution of X.

First, consider the following two examples to illustrate the eventual formula for the continuous case.

30

9.3.1 Example 1

Let $X \sim N(\mu, \sigma^2)$. Find the p.d.f. of $Z = \frac{X - \mu}{\sigma}$. We know the answer to this, but we don't know the proof for it.

Step 1: write down the c.d.f. of Z.

$$F_Z(z) = P(Z \leq z)$$
$$= P(\frac{X - \mu}{\sigma} \leq z)$$

Step 2: write the F_Z in terms of F_X.

$$P(\frac{X - \mu}{\sigma} \leq z) = P(X \leq \sigma z + \mu)$$
$$= F_X(\sigma z + \mu)$$

Step 3: differentiate in terms of z.

$$f_z(z) = \frac{d}{dz} F_Z(z)$$
$$= \sigma f_X(\sigma z + \mu) \quad \text{by chain rule}$$

Finally:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{1}{2}(\frac{x - \mu}{\sigma})^2} \Big|_{\sigma z + \mu}$$
$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad \forall z \in (-\infty, \infty)$$

9.3.2 Example 2 (Careful!)

Let $Z \sim N(0, 1)$. Find the p.d.f. of $Y = Z^2$.

$$F_Y(y) = P(Y \leq y)$$
$$= P(Z^2 \leq y)$$
$$= P(|Z| \leq \sqrt{y})$$
$$= P(-\sqrt{y} \leq Z \leq \sqrt{y})$$
$$= F_Z(\sqrt{y}) - F_Z(-\sqrt{y}) \quad \text{for } y \geq 0$$

Finally,

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy} F_Y(y) \\
&= \frac{1}{2} y^{-\frac{1}{2}} f_Z(\sqrt{y}) - \frac{1}{2} y^{-\frac{1}{2}} f_Z(-\sqrt{y}) \; \forall y \geq 0 \\
&= \frac{1}{2} y^{-\frac{1}{2}} (2 f_Z(\sqrt{y})) \; (N(0,1) \text{ density symmetric about } 0)
\end{aligned}
$$