Methods of Mathematical Statistics

Notes by Tim Brown and Guoqi Qian

Module 3: Continuous Distributions

Contents

1 Random Variables of the Continuous Type - 3.1

1.1 Cumulative Distribution Function (CDF), Probability Density Function (PDF)

Where have we been?

Discrete random variables have distinct values like $0, 1, \cdots$

Defined by their probability mass function, PMF, or moment generating function, MGF

Expectation, *cumulative probabilities, CDF* and variance can be computed by *sums* of PMF multiplied by function values (for example, 1 for cumulative probabilites, x^2 for expectation of square)

Sampling with and without replacement leads to Binomial and Hypergeometric distributions

Times till success lead to Geometric and Negative Binomial distributions

Poisson process arises from subdividng time in Bernoulli trials

Where are we going?

Continuous random variables can take any real value or an interval or a half-line of values

Defined by their probability *density* function, PDF, or moment generating function, MGF

Expectation, *cumulative probabilities, CDF* and variance can be computed by *integrals* of *PDF* multiplied by function values (for example, 1 for cumulative probabilites, x^2 for expectation of square)

PDF is found from CDF by *differentiation*

Normal distribution arises as limiting distribution of sample averages, including sample proportions

Start with the time to the first event in a Poisson process

This time can be *any* non-negative time - unlike the Geometric which can only be $0, 1, 2, \cdots$

Part 3 of Accidents in a Workplace Ex. - Discrete

Asked about the mean and standard deviation of time to 1st accident, T_1.

How do we find these for the Poisson process? Answer: need *PDF* and *CDF* - see Accidents example below.

Because the event that $T_1 > t$ $(t > 0)$ is the same as the event that $X_t = 0$, for a Poisson process, X_t with rate λ, the cumulative probability is

$$\begin{aligned} P(T_1 \leq t) &= 1 - P(T_1 > t) \\ &= 1 - P(X_t = 0) \\ &= 1 - e^{-\lambda t} \end{aligned} \tag{1}$$

New: as a function of t this is *continuous* - see Figure 1

Contrasts with the discrete random variable X_t where the graph of $P(X_t \leq x)$ has jumps - see Figures 2 and 3

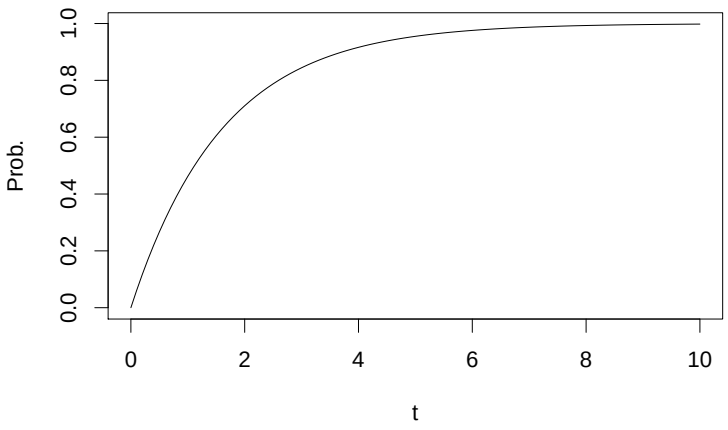

Figure 1: Cumualtive Probability $P(T_1 \leq t)$ for Poisson process

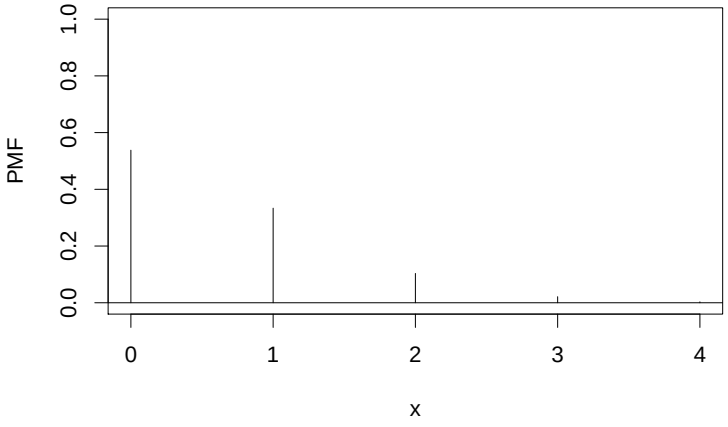

Figure 2: PMF for X_1 in Poisson process

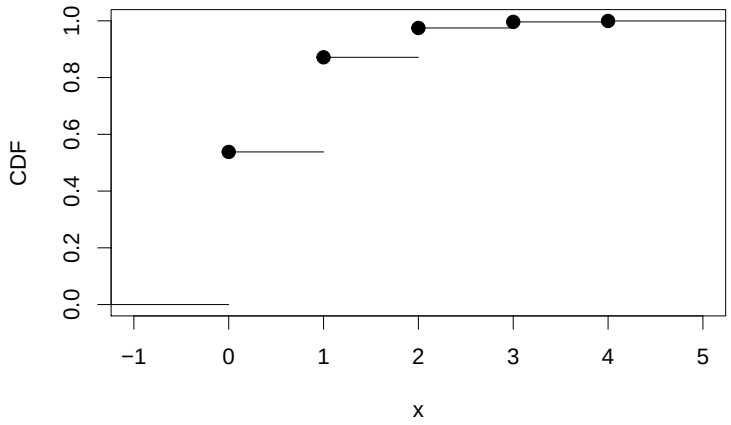

Figure 3: Cumulative Probability $P(X_1 \leq x)$ for Poisson process

Cumulative Distribution Function and Probability Density Function

Define for any continuous random variable X the *Cumulative Distribution Function, F, (CDF)* as the function defined for all real values x by

$$F(x) = P(X \leq x) \tag{2}$$

And call its derivative the *Probability Density Function, f, (PDF)* so

$$f(x) = F'(x) \tag{3}$$

Then Fundamental Theorem of Calculus gives

$$F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y \tag{4}$$

Density Function for T_1 in Poisson process

CDF previously derived:

$$\begin{aligned} F(x) &= P(T_1 \leq x) \\ &= \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases} \end{aligned}$$

PDF now obtained by differentiating

$$\begin{aligned} f(x) &= F'(x) \\ &= \begin{cases} 0, & x < 0 \\ \lambda e^{-\lambda x}, & x \geq 0 \end{cases} \end{aligned} \tag{5}$$

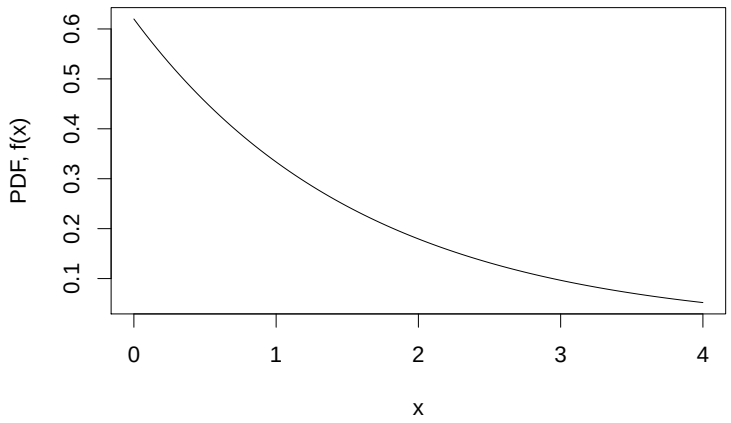

Figure 4: Probability Density Function, Called Exponential, for the RV T_1

Important Properties of Continuous RVs

CDF It is assumed that the CDF is differentiable except at "boundary" values (such as 0 in the Exponential case)

CDF is always increasing and has limit 0 towards $-\infty$ and limit 1 towards $+\infty$ because of properties of probability

PDF is non-negative and continuous (in this course) except at "boundary" values

PDF and CDF are defined for all real values x

Probs are *areas* under the PDF or *values* or differences of values of CDF - see example on next slide

Example - Accidents in a Workplace - Exponential

Suppose that accidents in the workplace occur at a rate of 0.62 per year. What is the probability that the first accident occurs between the end of the first year and the beginning of the third?

Solution - Accidents in a Workplace - Exponential

Let T_1 be the time to the 1st accident. Then from the fact that $[T_1 \leq 1]$ & $[1 < T_1 \leq 2]$ are disjoint with union $[T_1 \leq 2]$,

$$\begin{aligned} P(1 < T_1 \leq 2) &= P(T_1 \leq 2) - P(T_1 \leq 1) \\ &= (1 - e^{-0.62 \times 2}) - (1 - e^{-0.62 \times 1}) \\ &= 0.2485602 \end{aligned}$$

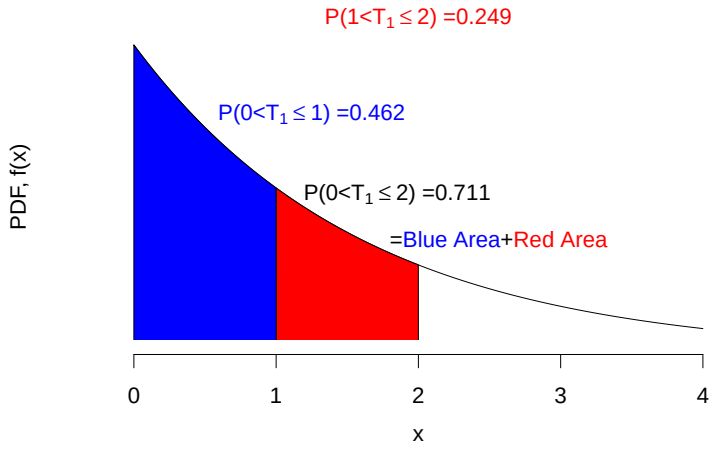

Figure 5: Exponential PDF - *red* area is desired probability

with the second line coming from equation 1. Figure 5 illustrates the areas under the pdf

1.2 MGF and Moments for Continuous Random Variables

Mean, Variance and MGF for Continuous Random Variables

Discrete RV X with pmf f and a real-valued function u has

$$E(u(X)) = \sum_{x \in \text{range} X} u(x) f(x).$$

Analogy for a continuous random variable X with probability density function f

$$E(u(X)) = \int_{-\infty}^{\infty} u(x) f(x)\, \mathrm{d}x. \tag{6}$$

Mean

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, \mathrm{d}x. \tag{7}$$

Mean, Variance and MGF for Continuous Random Variables

MGF

$$E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\, \mathrm{d}x. \tag{8}$$

Second Moment:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, \mathrm{d}x. \tag{9}$$

Recall that variance is the difference between the second moment and the square of the mean - still true for continuous random variables.

Rules the Same - *Just replace sums with integrals* .

2 The Exponential, Gamma and Chi-Square Distributions - 3.2

2.1 Exponential and Gamma Distributions

Example - Accidents in a Workplace - Gamma

Find the moment generating function, mean and standard deviation for the time to the (a) first and (b) fourth accidents, assuming they occur according to a Poisson process with rate 0.62 per year. Also find the probability density function for the time to the fourth accident.

Solution Exponential - Moment Generating Function for T_1

PDF is given in equation (3) so equation (8) gives the MGF, M of T_1 as

$$\begin{aligned} M(t) &= E(e^{tX}) \\ &= \int_0^\infty e^{tx} 0.62 e^{-0.62x} \, \mathrm{d}x \end{aligned}$$

since the density is 0 for $x < 0$.

PDF is given in equation (3) so equation (8) gives the MGF, M of T_1 as

$$\begin{aligned} M(t) &= E(e^{tX}) \\ &= \int_0^\infty e^{tx} 0.62 e^{-0.62x} \, \mathrm{d}x \end{aligned}$$

since the density is 0 for $x < 0$ and this

$$= \frac{0.62}{0.62 - t} \int_0^\infty (0.62 - t) e^{-(0.62-t)x} \, \mathrm{d}x.$$

PDF is given in equation (3) so equation (8) gives the MGF, M of T_1 by

$$\begin{aligned} M(t) &= E(e^{tX}) \\ &= \int_0^\infty e^{tx} 0.62 e^{-0.62x} \, \mathrm{d}x \end{aligned}$$

since the density is 0 for $x < 0$ and this

$$= \frac{0.62}{0.62 - t} \int_0^\infty (0.62 - t) e^{-(0.62-t)x} \, \mathrm{d}x.$$

The integrand is, provided $t < 0.62$, the exponential pdf with rate $0.62 - t$, so

$$M(t) = \frac{0.62}{0.62 - t} \times 1.$$

since pdf's integrate to 1 - *this trick will be used frequently!*

Note on Exponential PDF

Must integrate to 1 because

Limit of $P(T_1 \leq x)$ as $x \to \infty$ is 1 (check this if you like in equation (1)) so

$$\begin{aligned} 1 &= \lim_{x\to\infty} P(T \leq x) \\ &= \lim_{x\to\infty} \int_{-\infty}^{x} f(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \\ &= \int_{0}^{\infty} \lambda e^{-\lambda x}\,\mathrm{d}x \end{aligned}$$

Solution Exponential - Mean and SD for T_1

MGF derivatives, for $t < 0.62$ are:

$$M'(t) = \frac{0.62}{(0.62-t)^2}, \quad M''(t) = \frac{2 \times 0.62}{(0.62-t)^3}$$

Mean:

$$E(T_1) = M'(0) = \frac{1}{0.62} = 1.613$$

SD:

$$\begin{aligned} \sqrt{Var(T_1)} = \sqrt{M''(0) - (E(T_1))^2} &= \sqrt{\frac{2}{0.62^2} - (E(T_1))^2} \\ &= 1.613 \end{aligned}$$

Solution Gamma - MGF and Moments for T_4

Important: Just as in discrete time, the inter-accident times $T_1, T_2 - T_1, \cdots$ are *independent and identically distributed* - in continuous time, they all have the exponential distribution with rate the same as the rate of the Poisson process.

Proof is similar to discrete time. For example, if $t, s > 0$, arguing informally using the independence properties of the Poisson process:

$$\begin{aligned} P(T_1 > t \cap T_2 - T_1 > s) &= P(T_1 > t \cap X_{T_1+s} - X_{T_1} = 0) \\ &= P(T_1 > t) P(X_{T_1+s} - X_{T_1} = 0) \\ &= e^{-0.62t} e^{-0.62s}. \end{aligned}$$

Solution Gamma - MGF and Moments for T_4 ctd

So T_4 is the sum of 4 independent random variables each with the exponential distribution with rate = 0.62.

Hence using the fact that the MGF of a sum of independent random variables is the product of their MGF's, the MGF M_4 of T_4 is given by

$$M_4(t) = (M(t))^4 = \left(\frac{0.62}{0.62 - t}\right)^4$$

Expected value of any sum is the sum of the expectations, so

$$E(T_4) = 4E(T_1) = \frac{4}{0.62} = 6.451$$

Variance of a sum of *independent* random variables is the sum of the variances, so the standard deviation of T_4 is

$$\sqrt{Var(T_4)} = \sqrt{4Var(T_1)} = 3.226$$

Solution Gamma - PDF for T_4

CDF F_4 can be obtained from the argument used in the example at the end of Module 2, so for $t > 0$

$$\begin{aligned} F_4(t) &= 1 - P(T_4 > t) \\ &= 1 - P(X_t \leq 3) \\ &= 1 - \left(e^{-0.62t} + 0.62te^{-0.62t} \right. \\ &\quad \left. + \frac{(0.62t)^2 e^{-0.62t}}{2} + \frac{(0.62t)^3 e^{-0.62t}}{6}\right). \end{aligned}$$

Solution Gamma - PDF for T_4 ctd

PDF f_4 is obtained by differentiating - note the remarkable cancellation in the telescoping sum:

$$\begin{aligned} f_4(t) &= F_4'(t) \\ &= 0.62e^{-0.62t} - 0.62e^{-0.62t} + 0.62^2 te^{-0.62t} - 0.62^2 te^{-0.62t} \\ &\quad + \frac{0.62^3 t^2 e^{-0.62t}}{2} - \frac{0.62^3 t^2 e^{-0.62t}}{2} + \frac{0.62^4 t^3 e^{-0.62t}}{6} \\ &= \frac{0.62^4 t^3 e^{-0.62t}}{6}. \end{aligned}$$

In general, T_n is Σ of n independ't exp'l RVs rate λ:

MGF of the time, T_n to the nth event of a Poisson process of rate λ is M_n given, for $t < \lambda$ by

$$M_n(t) = \left(\frac{\lambda}{\lambda - t}\right)^n. \tag{10}$$

Mean & SD

$$E(T_n) = \frac{n}{\lambda}, \quad SD(T_n) = \frac{\sqrt{n}}{\lambda}. \tag{11}$$

PDF f_n is given by

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}, & x \geq 0. \end{cases} \tag{12}$$

Gamma is the name of the distribution of T_n.

Proofs

MGF & Moments: We found the MGF for T_1 with rate 0.62 in the example - general λ same. The MGF for T_n now follows, as in the example, by the fact that the inter-event times $T_i - T_{i-1}$ are independent each with the same exponential pdf as T_1. Differentiate and set to 0 for moments or use expectation and sum of *independent* rv's.

CDF With X_t a Poisson process of rate λ,

$$P(T_n \leq x) = 1 - P(X_t \leq n-1) = 1 - \sum_{i=0}^{n-1} \frac{e^{-\lambda}\lambda^i}{i!}$$

.

PDF Differentiate the CDF with respect to x and get the same telescoping sum as in the example - see text p. 106 for full write-up.

2.2 When to use Binomial, Geometric, Negative Binomial, Poisson, Exponential, Gamma

When to use?

Binomial, Geometric, Negative Binomial, Poisson, Exponential, Gamma				
Time	Assumptions	Counts	Inter-event Times	Event Times
Discrete	Bernoulli trials - independent, equal probabilities	Binomial	Independent Geometric	Negative Binomial
Continuous	No multiple points, independence of counts in disjoint time intervals, mean count = in = length intervals	Poisson	Independent Exponential	Gamma

2.3 General Gamma and Chi-square

More general Gamma distributions

Gamma function defined for any value $\alpha > 0$ by

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y}\, dy \tag{13}$$

PDF The probability density function for a general Gamma with *shape* parameter $\alpha > 0$ and *rate* parameter $\lambda > 0$ is given by

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, & x \geq 0 \end{cases} \tag{14}$$

Existence of Gamma Function

Gamma function for integer α is equal to $(\alpha - 1)!$ because, if f is the pdf of the time to the αth event of a Poisson process of rate *1*, then

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(y)\, dy \\ &= \int_0^{\infty} \frac{y^{\alpha-1} e^{-y}}{(\alpha-1)!}\, dy. \end{aligned}$$

So

$$\begin{aligned} (\alpha-1)! &= \int_0^{\infty} y^{\alpha-1} e^{-x}\, dy \\ &= \Gamma(\alpha), \quad \text{by the definition of } \Gamma(\alpha). \end{aligned} \tag{15}$$

Existence of Gamma Function Ctd

For non-integer $\alpha > 1$, let $[\alpha]$ be the integer part of α, that is the largest integer $< \alpha$. Then for $y > 0$

$$y^{\alpha-1} < \begin{cases} 1, & y \leq 1 \\ y^{[\alpha]}, & y > 1 \end{cases} \tag{16}$$

For $\alpha \leq 1$

$$y^{\alpha-1} e^{-y} < \begin{cases} y^{\alpha-1}, & y \leq 1 \\ y e^{-y}, & y > 1 \end{cases} \tag{17}$$

Existence of Gamma Function Ctd 2

Hence using (16) and (17)

$$\begin{aligned} \Gamma(\alpha) &= \int_0^{\infty} y^{\alpha-1} e^{-y}\, dy \\ &= \int_0^1 y^{\alpha-1} e^{-y}\, dy + \int_1^{\infty} y^{\alpha-1} e^{-y}\, dy \\ &< \begin{cases} \left[\frac{1}{\alpha} y^\alpha\right]_0^1 + \int_1^{\infty} y e^{-y}\, dy, & \alpha \leq 1 \\ 1 + \int_1^{\infty} y^{[\alpha]} e^{-y}\, dy, & \alpha > 1 \end{cases} \\ &< \begin{cases} \frac{1}{\alpha} + 1, & \alpha \leq 1 \\ 1 + ([\alpha]+1)!, & \alpha > 1 \end{cases} \qquad < \infty \end{aligned}$$

with the 1 in the case $\alpha \leq 1$ coming from the fact that the integral from 1 to ∞ is bounded by the mean of an exponential pdf with rate 1 and the factorial from (15)

Existence of Gamma PDF

Now that we know that the Gamma function evaluated at *any* $\alpha > 0$ is a finite number, can show that the pdf in (14) integrates to on

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \frac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}\, dx = \int_{0}^{\infty} \frac{y^{\alpha-1} e^{-y}}{\Gamma(\alpha)}\, dy$$

on letting $y = \lambda x$ so that $dy' =' \lambda dx$ (in the sense that the derivative of y as a function of x is the constant λ).

So

$$\int_{-\infty}^{\infty} f(x)dx = \frac{1}{\Gamma(\alpha)} \int_{0}^{\infty} y^{\alpha-1} e^{-y}\, dy = 1,$$

as required, by the definition of the Gamma function.

General Gamma Ctd

Rate or Scale? To show that the PDF integrates to one for general λ and not just the case $\lambda = 1$, a change of variables in the integral is necessary. The number $\theta = 1/\lambda$ is more convenient in this calculation and is thus called the *scale* of the Gamma distribution. The PDF becomes

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{x^{\alpha-1} e^{-x/\theta}}{\theta^{\alpha}\Gamma(\alpha)}, & x \geq 0 \end{cases} \tag{18}$$

PDF& CDF pictures for varying shape and scale parameters are given in Figures 6, 7, 8 and 9. For *non-integer shape*, the CDF does not have an easy closed form just like the Gamma function for non-integers does not have an easy closed form.

MGF & Moments

Follow as previously since it can be shown that the independence properties continue to hold with non-integer shape.

Scale is the reciprocal of *rate* so for $t < 1/\theta$ the moment generating function, M, for a Gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\theta > 0$ is

$$M(t) = \frac{1}{(1-\theta t)^{\alpha}} \tag{19}$$

Mean & SD The mean, μ, and standard deviation, σ, for the Gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\theta > 0$ are

$$\mu = \alpha\theta, \quad \sigma = \sqrt{\alpha}\theta \tag{20}$$

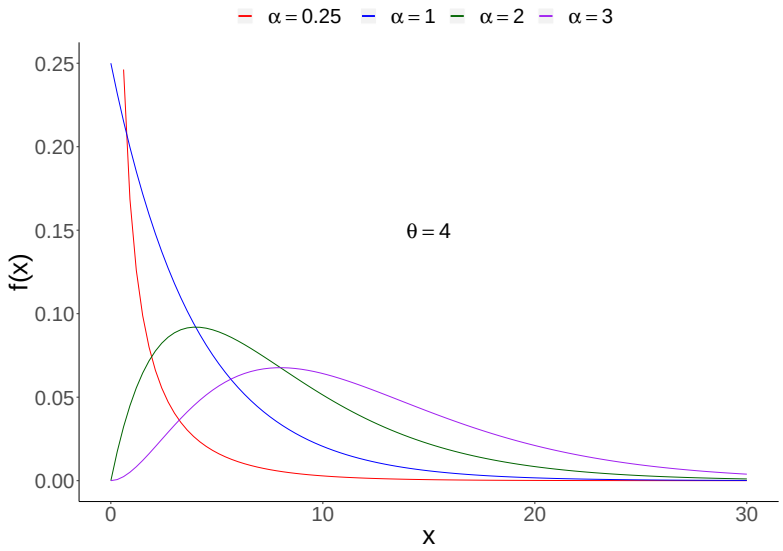

Figure 6: Gamma Probability Density Functions - Varying Shape

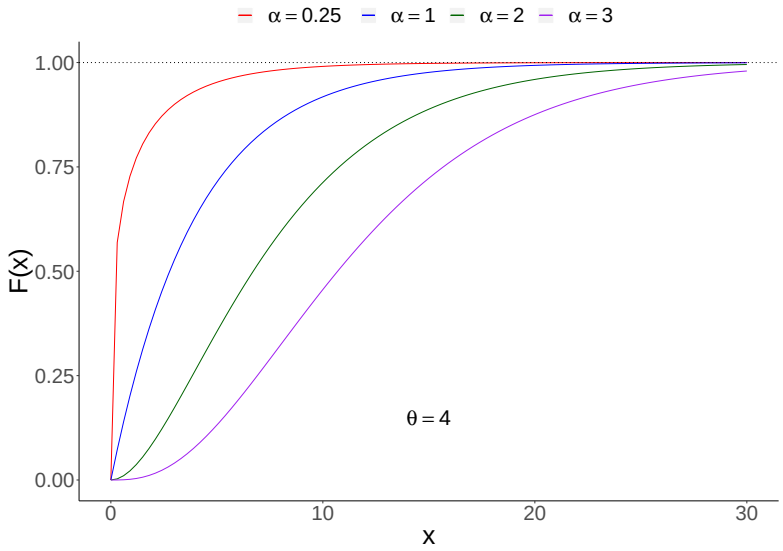

Figure 7: Gamma Cumulative Distribution Functions - Varying Shape

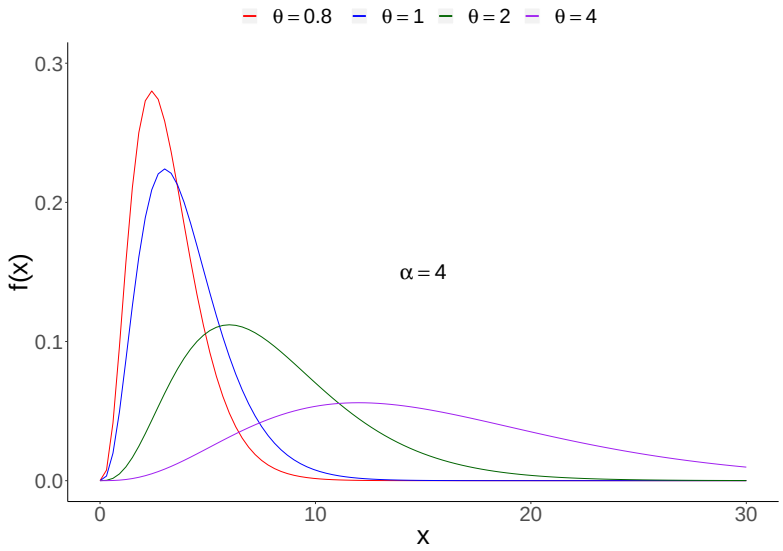

Figure 8: Gamma Probabiliity Density Functions - Varying Scale

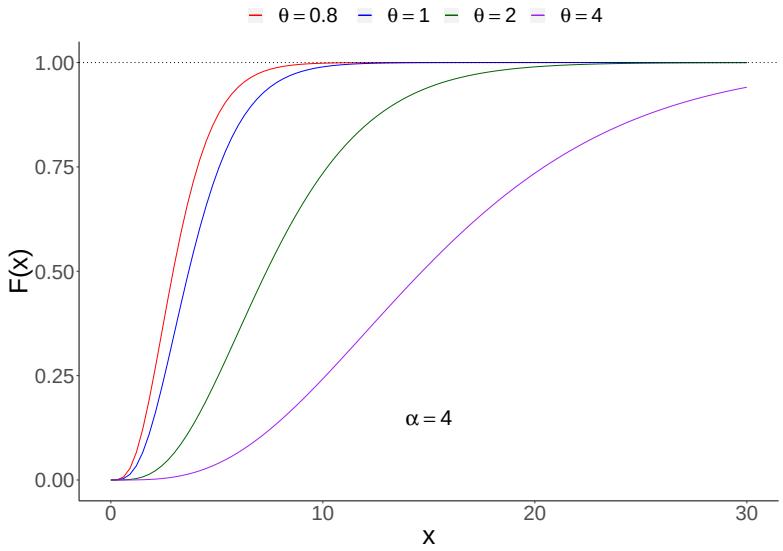

Figure 9: Gamma Cumulative Distribution Functions - Varying Scale

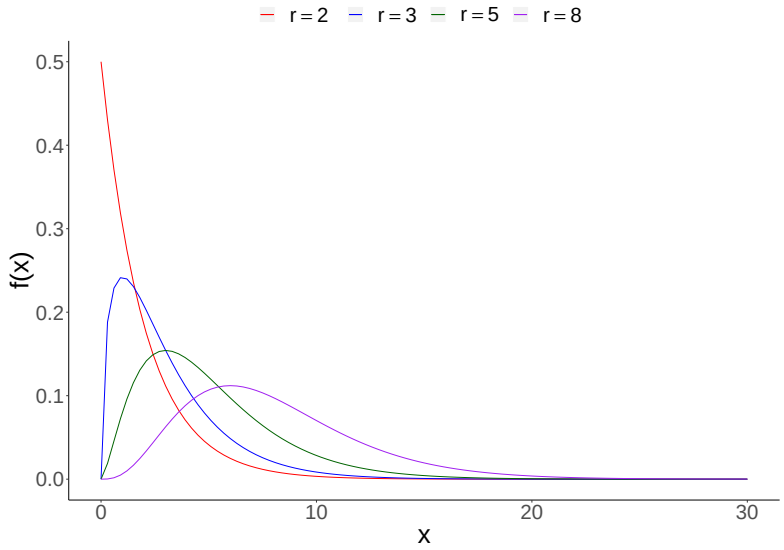

Figure 10: Chi-Square PDFs - Varying Degrees of Freedom

Chi-Square

The Gamma density with shape being a multiple of $\frac{1}{2}$, say $\alpha = \frac{r}{2}$, for some integer $r = 0, 1, 2, \cdots$ and scale *2* is called the Chi- square distribution and denoted $\chi^2(r)$. The pdf is:

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{x^{r/2-1}e^{-x/2}}{2^{r/2}\Gamma(r/2)}, & x \geq 0 \end{cases} \tag{21}$$

Degrees of Freedom: - name given to the parameter r for the Chi-square distribution. There are many statistical applications of the Chi-square distribution. In these applications, the name *degrees of freedom* is natural. Figures 10 and 11 show the Chi-square pdf and cdf for varying degrees of freedom.

Chi-Square Distribution

MGF For $t < 2$ the MGF, M, for the $\chi^2(r)$ distribution is:

$$M(t) = \frac{1}{(1-2t)^{r/2}}. \tag{22}$$

Mean & SD The mean, μ, and standard deviation, σ, for the $\chi^2(r)$ distribution are:

$$\mu = r, \quad \sigma = \sqrt{2r}. \tag{23}$$

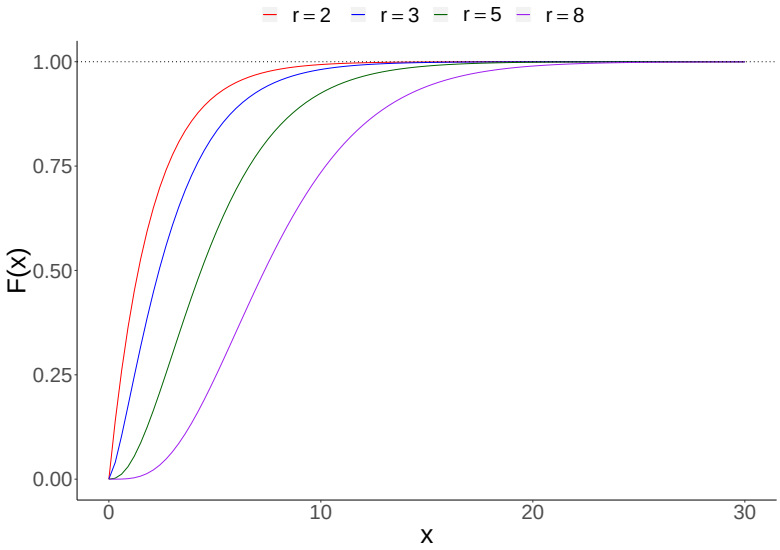

Figure 11: Chi-Square CDFs - Varying Degrees of Freedom

2.4 Quantiles and Percentiles - 6.3

Quantiles - 6.3

Values from the inverse of the CDF are called *quantiles* of the distribution. Formally, the pth quantile for the CDF, π_p, is the value solving

$$F(\pi_p) = p. \tag{24}$$

Discrete distributions may have no solutions to equation (24) and at values in the range of the random variable there will be infinitely many distributions, because the CDF is flat apart from jumps at the points of the range. Usually the left-hand end point of a flat part is taken as the percentile.

Continuous distributions have only one solution - provided their density has no flat spots ensuring that, for each $0 < p < 1$, π_p is uniquely defined.

Percentiles

Percentiles

Percentiles The $100p$th percentile is the pth quantile

Median is the 50th percentile

Example - Chi-Square

Suppose a random variable X has a chi-square distribution with two degrees of freedom. What is its probability density function? Does it have another name? What are the mean and median, as well as the 5th, 25th, 75th and 95th percentiles for the random variable X?

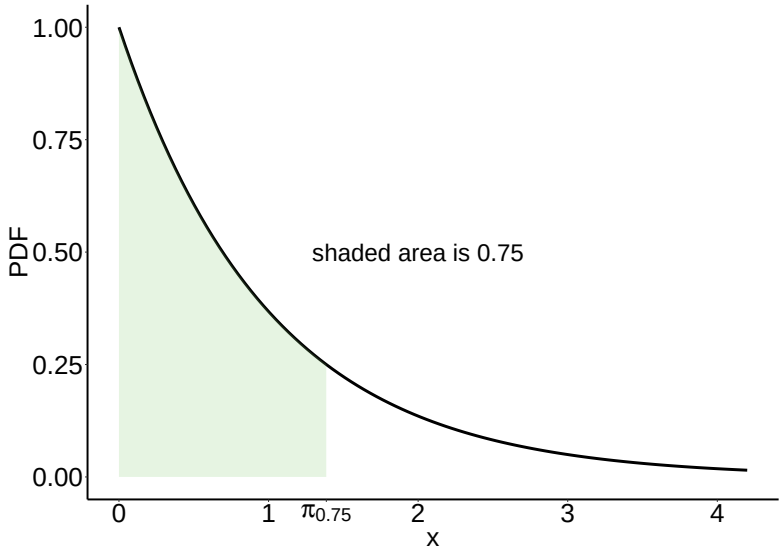

Figure 12: 0.75 Quantile is $\pi_{0.75}$ - PDF graph

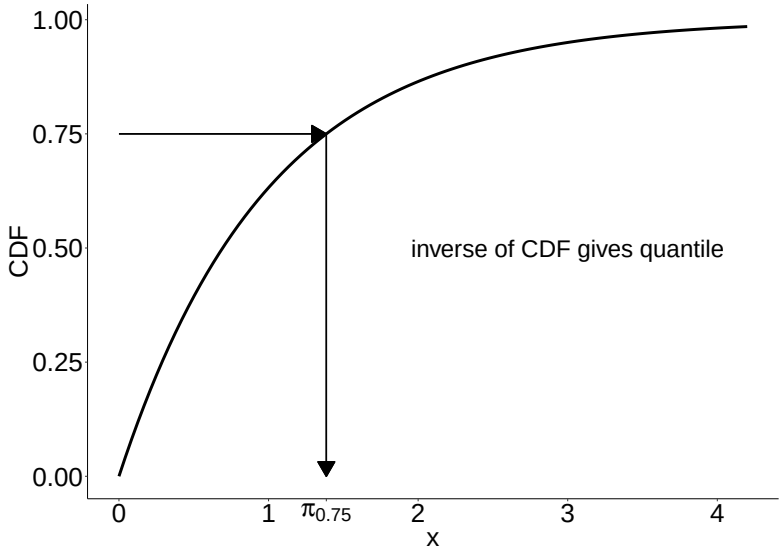

Figure 13: 0.75 Quantile is $\pi_{0.75}$ - CDF graph

Solution - Chi-Square

PDF is:

$$f(x) = \begin{cases} 0, & x < 0 \\ \frac{x^{2/2-1}e^{-x/2}}{2^{2/2}\Gamma(2/2)}, & x \geq 0 \end{cases}$$
$$= \begin{cases} 0, & x < 0 \\ \frac{e^{-x/2}}{2}, & x \geq 0 \end{cases}$$

which is the pdf of exponential with rate $1/2$, or equivalently scale 2!

The mean is 2 and the quantiles, π_p solve

$$F(\pi_p) = 1 - e^{-1/2\pi_p} = p.$$

So

$$\pi_p = -2\log(1-p).$$

Solution - Chi-Square Ctd

Median is $-2\log(0.5) = 1.39$ - less than the mean because of the skewness.

5th 25th,75th and 95th percentiles are $-2\log(1-p)$ for $p = 0.05, 0.25, 0.75, 0.95$ so they are
0.013, 0.575, 2.773, 5.991 (respectively).

3 Uniform Distribution, Random Numbers, CDF etc from PDF - 3.1, 5.1

3.1 Random Numbers

Example - Random Numbers

Random numbers generated by a computer between 2 and 4 are all equally likely. What is the probability density for such random numbers? Sketch it. What is the mean and standard deviation of the numbers? If 1,000,000 are generated, what should the mean and standard deviation of the generated numbers approximate?

Solution - Random Numbers

Model the random numbers as coming from a continuous distribution as this will cover all architectures and software generating the random numbers.

Assumption means that the density must be constant on $[2, 4]$ to give equally likely outcomes there.

Let f be the density with value c (say) between 2 and 4. Then

$$
\begin{aligned}
1 &= \int_{-\infty}^{\infty} f(x)\,dx \\
&= \int_{2}^{4} c\,dx \\
&= \Big[cx\Big]_{2}^{4} \\
&= 2c
\end{aligned}
$$

so $c = {}^{1}\!/_{2}$.

Solution - Random Numbers Ctd

Hence

$$
f(x) = \begin{cases} 0, & x < 2, x > 4 \\ \frac{1}{2}, & 2 \le x \le 4 \end{cases}
$$

Figure 14 shows the PDF of the random numbers.

Mean If μ is the mean of the pdf f, then

$$
\begin{aligned}
\mu &= \int_{-\infty}^{\infty} x f(x)\,dx \\
&= \int_{2}^{4} \frac{x}{2}\,dx \\
&= \left[\frac{x^2}{4}\right]_{2}^{4} \\
&= 3.
\end{aligned}
$$

Solution - Random Numbers Ctd 2

Variance: If σ is the SD for the pdf f, then

$$
\begin{aligned}
\sigma^2 &= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2 \\
&= \int_{2}^{4} \frac{x^2}{2}\,dx - 9 \\
&= \left[\frac{x^3}{6}\right]_{2}^{4} - 9 \\
&= \frac{1}{3}.
\end{aligned}
$$

Hence

$$
\sigma = \frac{1}{\sqrt{3}} = 0.5773
$$

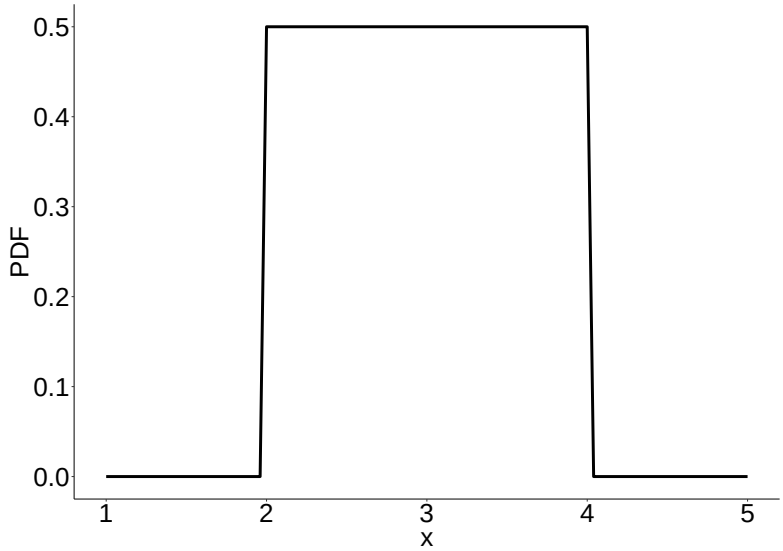

Figure 14: PDF of the Random Numbers

Solution - Random Numbers Ctd 3

If 1,000,000 random numbers are generated, the mean and sd of the 1,000,000 numbers would be expected to be close to μ and σ.

How close? - see the subsection in this Module on the Central Limit Theorem

Example - CDF from the PDF

Find the CDF from the PDF of the uniform random numbers and plot it

Solution - CDF from the PDF

Let F be the CDF of the random numbers. For $2 \leq x \leq 4$

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(y)\, dy \\ &= \int_{2}^{x} \frac{1}{2}\, dy \\ &= \left[\frac{y}{2}\right]_{2}^{x} \\ &= \frac{x}{2} - 1. \end{aligned}$$

For $x < 2$, $F(x) = 0$. For $x > 4$, $F(x) = 1$.

Solution - CDF from the PDF Ctd

Summarising:

$$F(x) = \begin{cases} 0, & x < 2 \\ \frac{x}{2} - 1, & 2 \leq x \leq 4 \\ 1, & x > 4 \end{cases}$$

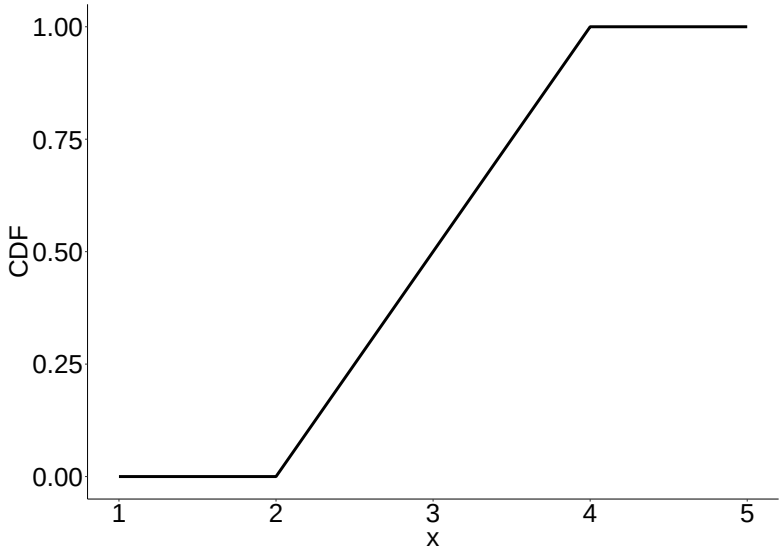

Figure 15: CDF of Random Numbers

Figure 15 shows this CDF

Example - Other Random Numbers

If U is a random variable with PDF which is 1 between 0 and 1, and 0 elsewhere, what is the CDF of 2U+2? If a computer only produces equally likely random numbers between 0 and 1, how would you produce equally likely numbers between 2 and 4?

Solution - Other Random Numbers

The PDF of U is f given by

$$f(x) = \begin{cases} 0, & x < 0, x > 1 \\ 1, & 0 \leq x \leq 1. \end{cases}$$

Figure 16 shows the relationship of events for $2U + 2$ and U. The CDF, F, of $2U + 2$ is given, for $2 < x < 4$ by

$$\begin{aligned} F(x) &= P(2U + 2 \leq x) \\ &= P(U \leq \frac{x-2}{2}) \\ &= \int_0^{\frac{x-2}{2}} 1 \, dy \\ &= \Big[y \Big]_0^{\frac{x-2}{2}} = \frac{x}{2} - 1. \end{aligned}$$

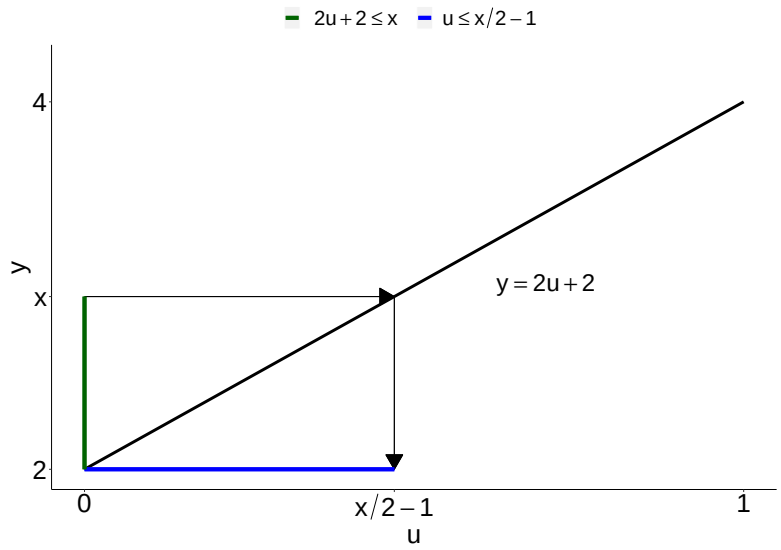

Figure 16: Event $[2U + 2 \leq x] = [U \leq \frac{x}{2} - 1]$

Solution - Other Random Numbers Ctd

Same as the CDF for the random numbers between 2 and 4.

So random numbers from 0 to 1 can be turned into random numbers between 2 and 4 by multiplying them by 2 and then adding 2.

Comment: All random numbers are based on the equally likely ones between 0 and 1.

Algorithms: use generalisations of the argument we used here

More in Ch 5.1 and 5.2, as well as 3.3 on the Normal Distribution standardized random variable and square of it.

Uniform Distribution

Name of the distribution of U in the last example is the *Uniform Distribution on* $[0, 1]$.

Name of the distribution for the equally likely random numbers between 2 and 4 is the *Uniform Distribution on* $[2, 4]$.

PDF CDF, Mean, SD for general Uniform distribution on any interval are found by the same arguments we used in the example on random numbers on $[2, 4]$.

4 Normal Distribution - 3.3

4.1 Definition

PDF

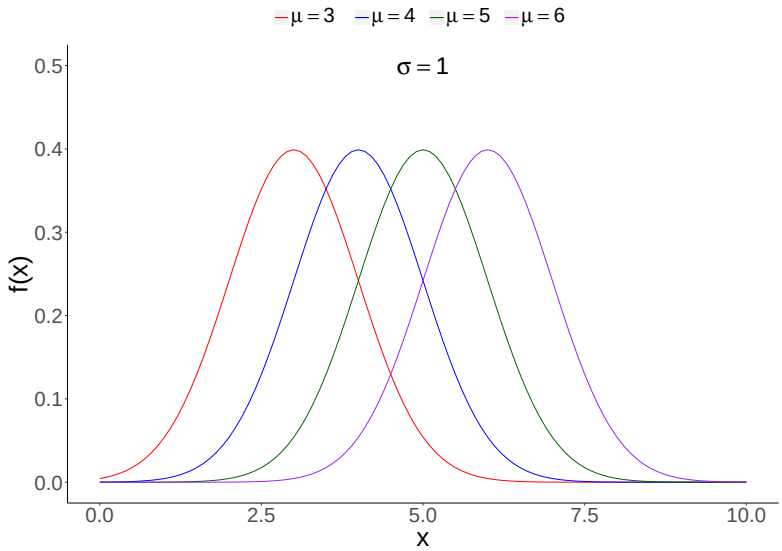

Figure 17: Normal PDFs - Varying μ

Definition: The PDF of a Normal Distribution with parameters μ and σ is the function f given, for any real x by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \tag{25}$$

Figures 17 and 18 show the PDF for varying values of μ and σ.

Notation: a random variable X with density function f is written as $X \sim N(\mu, \sigma^2)$.

CDF

CDF cannot be expressed as a finite sum of elementary functions like polynomials or exponentials or logs or trignometric funtions.

Showing that the pdf integrates to 1 requires a trick by squaring the integral, rewriting the square as a double integral and then transforming to polar co-ordinates.

Figures 19 and 20 show normal CDFs for varying μ and σ.

Normal probabilities are found from packages like R or Mathematica, or historically from tables.

4.2 MGF, Mean and SD, Percentiles

MGF

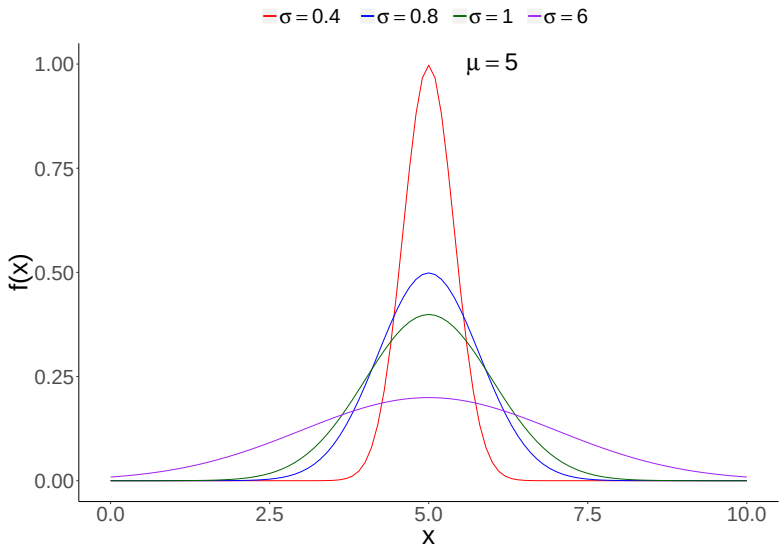

Figure 18: Normal PDFs - Varying σ

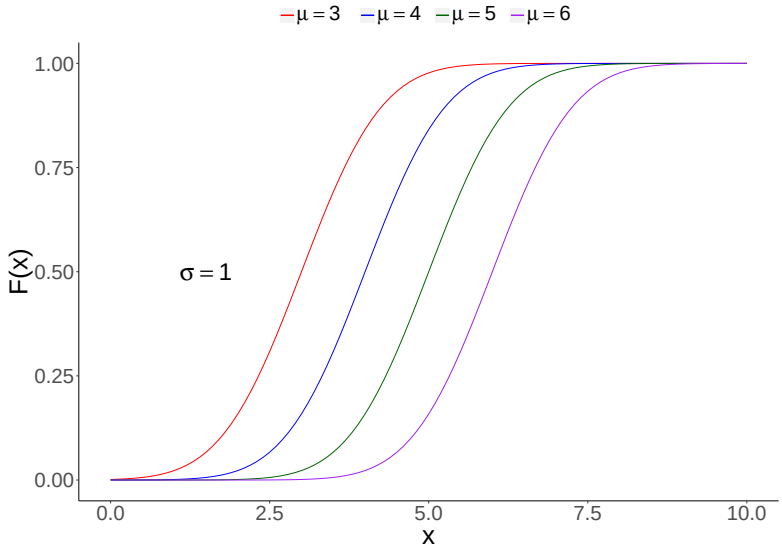

Figure 19: Normal CDFs - Varying μ

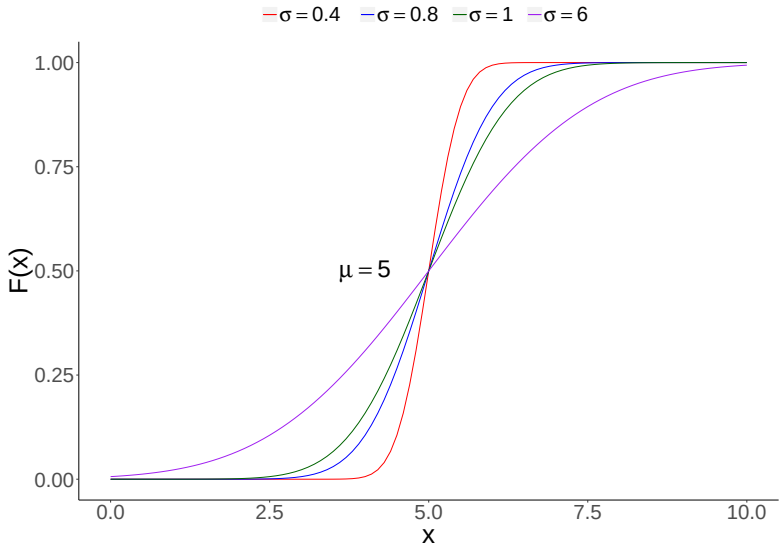

Figure 20: Normal CDFs - Varying σ

Suppose $X \sim N(\mu, \sigma^2)$ and that M is the moment generating function of X, so for any real t,

$$\begin{aligned} M(t) &= E(e^{tX}) \\ &= \int_{-\infty}^{\infty} \exp[tx] \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx. \end{aligned}$$

Multiplying the two exp'ls gives the exponential of the following sum:

$$\begin{aligned} tx - \frac{(x-\mu)^2}{2\sigma^2} &= -\frac{1}{2\sigma^2}\left[-2tx\sigma^2 + (x-\mu)^2\right] \\ &= -\frac{1}{2\sigma^2}\left[(x-(\mu+t\sigma^2))^2 - 2\mu\sigma^2 t - \sigma^4 t^2\right], \end{aligned}$$

on expanding $(x-\mu)^2$ and completing the square.

MGF Ctd

Hence

$$\begin{aligned} M(t) = \exp\left[\frac{2\mu\sigma^2 t + \sigma^4 t^2}{2\sigma^2}\right] \times \\ \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}\right] dx. \end{aligned}$$

But the integral is that of the $N(\mu+\sigma^2 t, \sigma^2)$ pdf so $= 1$. Hence

$$\begin{aligned} M(t) &= \exp\left[\frac{2\mu\sigma^2 t + \sigma^4 t^2}{2\sigma^2}\right] \\ &= \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right]. \end{aligned} \tag{26}$$

Derivatives of MGF

First:

$$M'(t) = (\mu + \sigma^2 t) \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right] = (\mu + \sigma^2 t) M(t)$$

Second: use the product rule to get

$$\begin{aligned} M''(t) &= \left(\sigma^2 M(t) + (\mu + \sigma^2 t) M'(t)\right) \\ &= \left((\mu + \sigma^2 t)^2 + \sigma^2\right) M(t) \end{aligned}$$

Moments

Mean: If $X \sim N(\mu, \sigma^2)$, then

$$\begin{aligned} E(X) &= M'(0) \\ &= (\mu + \sigma^2 \times 0) M(0) = \mu \end{aligned} \tag{27}$$

Variance:

$$\begin{aligned} Var(X) &= M''(0) - \mu^2 \\ &= \left((\mu + \sigma^2 \times 0)^2 + \sigma^2\right) M(0) - \mu^2 \\ &= \sigma^2 \end{aligned} \tag{28}$$

SD:

$$SD(X) = \sqrt{Var(X)} = \sigma \tag{29}$$

Example - Normal calculations

If $X \sim N(3, 1)$, what is the moment generating function, mean and standard deviation of X? What are the median and the 5th, 25th, 75th and 95th percentiles for X?

Solution - Normal Calculations

MGF: from equation (26) the MGF is

$$M(t) = \exp\left[3t + \frac{t^2}{2}\right]$$

Mean & SD: from equations (27) and (29)

$$E(X) = 3; SD(X) = 1$$

Percentiles other than the median - which is 3 by symmetry of the pdf - need to come from R or Mathematica and here are the R commands to give the required percentiles to 2 dp:

```
options(digits = 3)
p <- c(0.05, 0.25, 0.75, 0.95)
qnorm(mean = 3, sd = 1, p)

## [1] 1.36 2.33 3.67 4.64
```

4.3 Standardised Random Variable, Z

Definition of Z

Suppose $Z \sim N(0,1)$. Then $E(Z) = 0, Var(Z) = 1$ and Z is called a *standard* normal random variable. Suppose $\sigma > 0$, μ is any number and $X = \sigma Z + \mu$. Then

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(\sigma Z + \mu \leq x) \\ &= P\left(Z \leq \frac{x-\mu}{\sigma}\right) \\ &= F_Z\left(\frac{x-\mu}{\sigma}\right). \end{aligned} \tag{30}$$

Differentiating with respect to x,

$$\begin{aligned} f_X(x) &= \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \end{aligned} \tag{31}$$

Z to X

So, since this is a normal density, $X \sim N(\mu, \sigma^2)$. Hence normal probabilities for X can be found from those for Z using equation (30). Or the number x is turned into *statistical units* as $\frac{x-\mu}{\sigma}$ and the standard normal CDF is applied to the number in statistical units. $100p$th percentile, π_p, for Z gives the $100p$th percentile for X as $\sigma\pi_p + \mu$.

4.4 Normal and Chi-Square

Normal & Chi-Square

Let $Z \sim N(0,1)$. Then the moment generating function of Z^2 is given by

$$\begin{aligned} E\left(\exp[Z^2 t]\right) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[z^2 t\right] \exp\left[-\frac{z^2}{2}\right] dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2(1-2t)}{2}\right] dz \\ &= (1-2t)^{-1/2} \times \\ &\int_{-\infty}^{\infty} \frac{(1-2t)^{1/2}}{\sqrt{2\pi}} \exp\left[-\frac{z^2(1-2t)}{2}\right] dz \end{aligned} \tag{32}$$

Integrand on the right hand side of (32) is the pdf of $N\left(0, \frac{1}{1-2t}\right)$ so integrates to 1. Thus the factor before the integral is the MGF of Z^2 and it is the MGF of χ^2_1, so $Z^2 \sim \chi^2_1$.

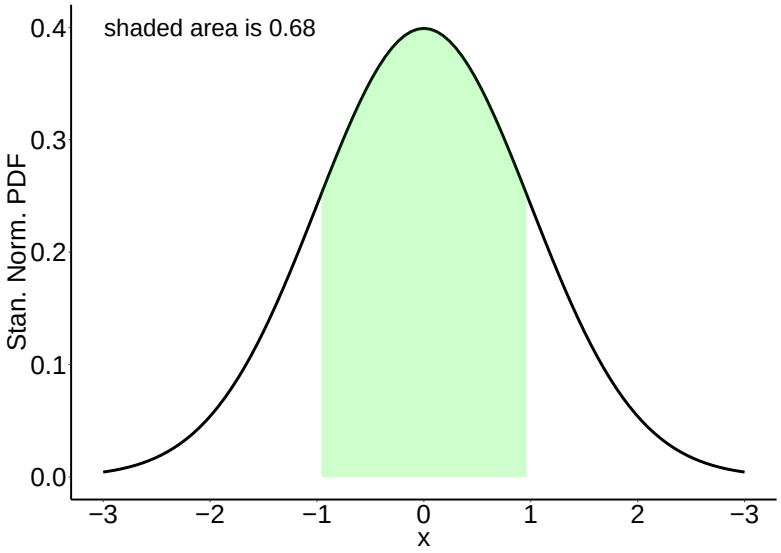

Figure 21: Standard Normal Distribution - 1 standard deviation from mean

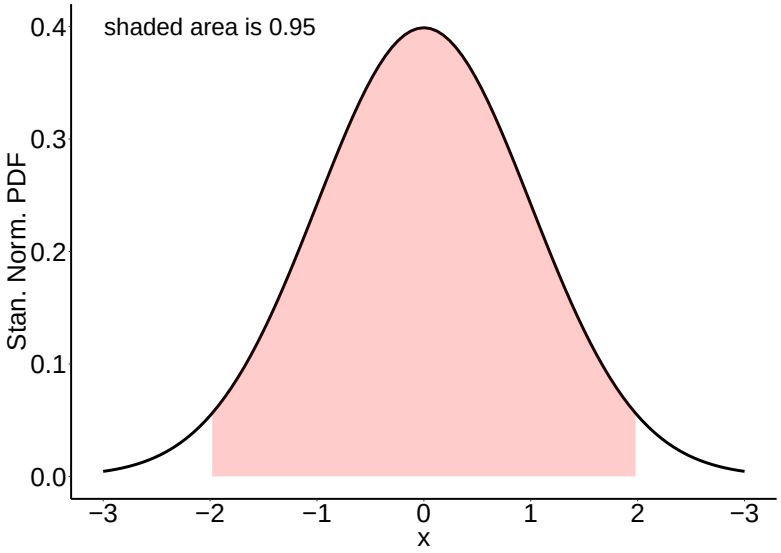

Figure 22: Standard Normal Distribution - 2 standard deviations from mean

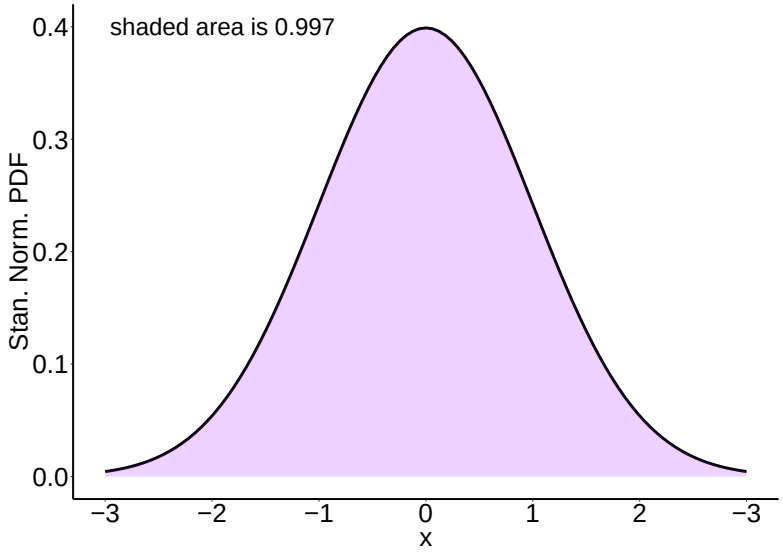

Figure 23: Standard Normal Distribution - 3 standard deviations from mean

Example - Errors in Measurement

A sensing device for location has normally distributed errors in each of the three directions that are independent and have mean 0 and standard deviation 0.1 m. What is the chance that the distance of the measured location is more than 0.3 m from the true location in 3D?

Solution - Errors in Measurement

Let $D_i \sim N(0, 0.1^2)$ be the three independent errors in measurement in height and planar locations

Then the distance of the measured location from the true location is $D = \sqrt{D_1^2 + D_2^2 + D_3^2}$, so the required probability is

$$P(D > 0.3) = P\left(\left[\frac{D}{0.1}\right]^2 > \left[\frac{0.3}{0.1}\right]^2 = 9\right)$$

since the two events are the same because the random variable D is non-negative - see Figure 24

Now, letting $Z_i = \frac{D_i}{0.1}$, $i = 1, 2, 3$ be the three independent standardised random variables,

$$\left[\frac{D}{0.1}\right]^2 = Z_1^2 + Z_2^2 + Z_3^2$$

Solution - Errors in Measurement Ctd

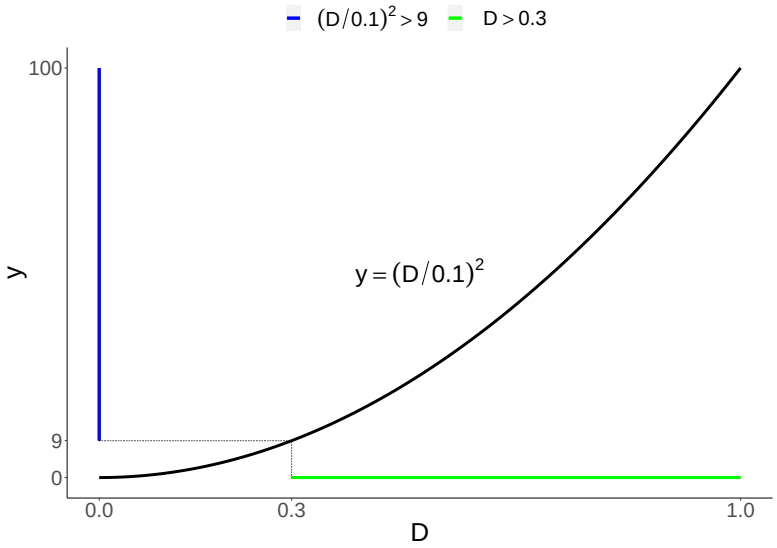

Figure 24: Event $[D > 0.3]$ is the same as $\left[\left[\frac{D}{0.1}\right]^2 > 9\right]$

MGF of a sum is the product of the MGF's and $Z_i^2 \sim \chi_1^2$ so the MGF $M(t)$ of the sum of these three is given by

$$M(t) = (1 - 2t)^{-3/2}$$

which is that of χ_3^2.

Hence required probability can be found from R using the commands with Figure 25 showing 1 -CDF for the total:

```
options(digits = 3)
1 - pchisq(9, df = 3)  # equivalently

## [1] 0.0293

1 - pgamma(9, shape = 3/2, scale = 2)

## [1] 0.0293
```

There is about 32% chance of one error being more than 0.1, but only 3% chance of the total error being $> 3 \times 0.1$.

Notes - Errors in Measurement Example

MGF, $M(t)$, of the sum of squares of n independent standard normals is the product of the MGF's and $Z_i^2 \sim \chi_1^2$ so

$$M(t) = (1 - 2t)^{-n/2}$$

which is that of χ_n^2.

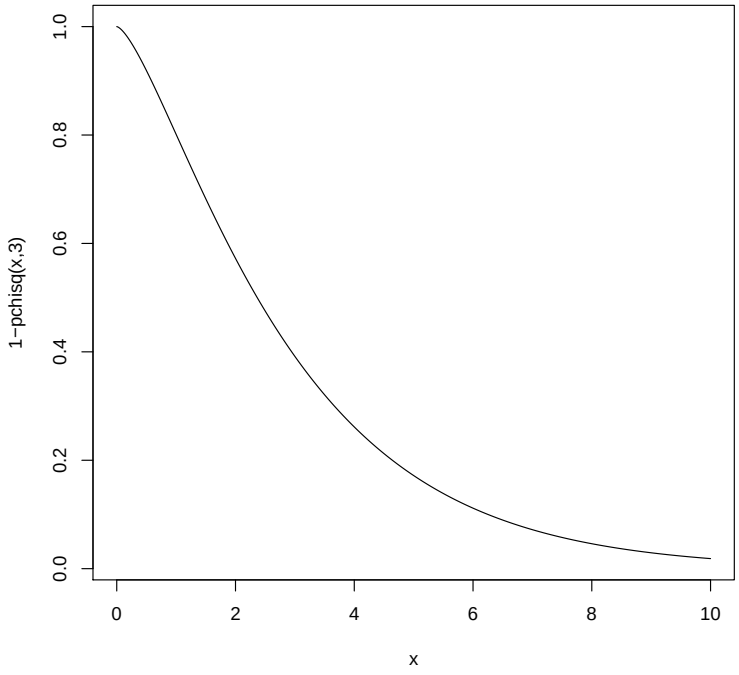

Figure 25: 1 - CDF of Chisq(3)

Think of the errors as each having 1 degree of freedom, so that their squares also each have 1 degree of freedom.

The sum of squares then has n degrees of freedom.

4.5 Central Limit Theorem and Law of Large Numbers - Ch 5.6 & 5.7

Central Limit Theorem

Suppose $X_1, X_2, \cdots X_n$ are independent random variables all with the same *arbitrary* distribution having mean μ and standard deviation σ.

Then letting $\bar{X}$ be the sample mean - $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$. From p.30 Module 2,

$$E(\bar{X}) = \mu; \quad SD(\bar{X}) = \frac{\sigma}{\sqrt{n}}. \tag{33}$$

CLT: Central Limit Theorem says that for any number z

$$P(\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \leq z) \to P(Z \leq z) \tag{34}$$

as $n \to \infty$, where $Z \sim N(0, 1)$.

Central Limit Theorem Consequences

If the the number of observations size is large, then the distribution of the mean of independent random variables with the same distribution is approximately normal.

The approximate normal distribution is centred at the true population mean and has standard deviation proportional to both the population standard deviation and the inverse of the square root of the number of observations.

So however small is the number $\epsilon > 0$

$$P(|\bar{X} - \mu| > \epsilon) = P(\frac{\sqrt{n}|\bar{X} - \mu|}{\sigma} > \frac{\sqrt{n}\epsilon}{\sigma}).$$

For large enough n, the right hand side is close to the normal probability of a large number and so arbitrarily small. Hence, $P(|\bar{X} - \mu| > \epsilon) \to 0, n \to \infty$. This is the *Law of Large Numbers* - the sample mean, $\bar{X}$, is arbitrarily close to the population mean, μ, with probability converging to one as the sample size increases.

Example - Conclusions in Opinion Polls

The Australian of 3rd April 2017 reported an opinion poll of 1708 voters showed that (primary) support for the Government had slipped from the previous week's poll of 37% to 36% this week. The headline was "Coalition in poll slip after tax win". Is this a fair headline assuming the support was actually exactly 37% in the whole population in the previous week?

Solution - Conclusions in Opinion Polls

Let $X_1, X_2, \cdots X_{1708}$ record 0 or 1 (respectively) according to whether the 1st, 2nd, ... , 1708th voter in the sample would not/ would (respectively) vote Liberal.

Then $X = \sum_{i=1}^{1708} X_i$ has a Binomial distribution (assuming sampling with replacement this is exact and otherwise nearly exact from the work done on the hypergeomtric distribution).

Assume the population proportion voting Liberal is 37%. Then the Binomial distribution for X has $n = 1,708$ and $p = 0.37$ so the probability that informs out judgement is $P(X \leq 0.36{*}1708)$. This can be found exactly from the R command:

```
pbinom(0.36 * 1708, size = 1708, prob = 0.37)

## [1] 0.191
```

Solution - Conclusions in Opinion Polls

There is 20% chance of getting a sample result of 36% or less with this sample size - so the headline seems unfair as it could just be a sample fluctuation.

CLT approximation: Since $Var(X_i) = 0.37 \times 0.63$, the Central Limit Theorem gives:

$$\frac{\sqrt{1708}(\bar{X} - 0.37)}{\sqrt{0.37 \times 0.63}} \approx N(0, 1)$$

And the R command shows that the approximation is close:

```
pnorm(-0.01 * sqrt(1708/(0.37 * 0.63)))

## [1] 0.196
```

The sample fluctuation explanation is even more plausible since the previous 37% result came from another independent poll, giving more variability in the difference - this will be discussed further in Modules 8 and 9.