Mathematical Background

A quick refresher of terms from calculus that will help in this course. Also, a review of some statistical terminology and definitions.

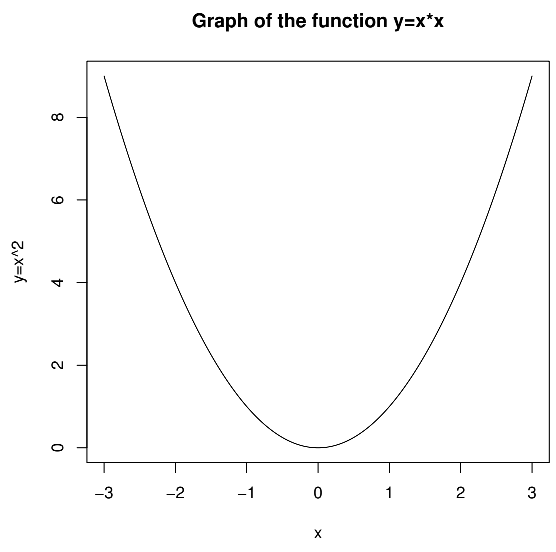
> Note: The following are *very* non-rigorous definitions designed to suit the purpose of our course. Refer to any calculus and/or statistics textbook for the exact definitions and/or more information.

Functions: For our purposes, a *function* assigns a unique numerical value to each number in a specified set. For example, the function

$$f(x) = x^2, \ -\infty < x < +\infty$$

assigns the value x^2 to each x, $-\infty < x < +\infty$. Thus $x = 1$ is assigned the value 1, $x = 2$ is assigned the value 4, and $x = -2.1$ is assigned the value $+4.41$, etc. A function is defined over a set of values, which here is the set of all real numbers.

Functions are often easily understood by looking at the *graph* of the function.

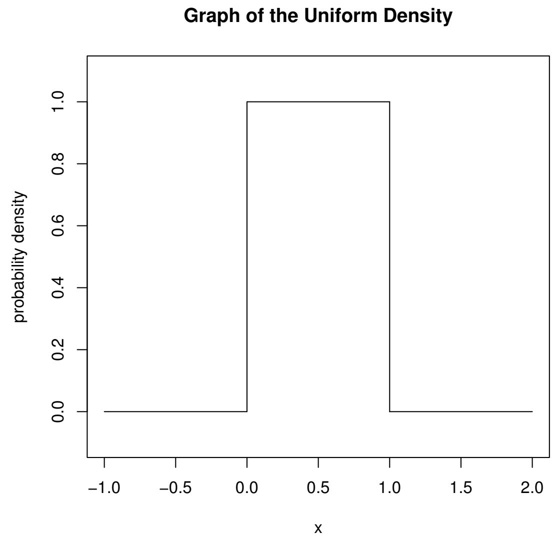


Functions are used in statistics to describe probability (density) functions (among many other things). Some examples:

(i) The Uniform probability (density) function describes the experiment of choosing a random number between 0 and 1. The function is

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise,} \end{cases}$$

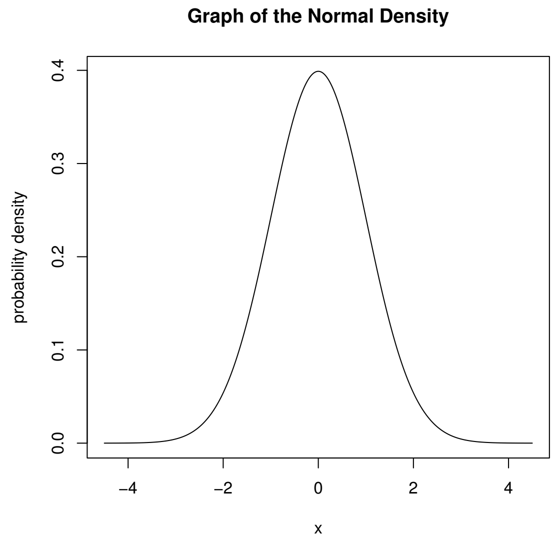
and the graph is shown below:



(ii) The standard Normal probability (density) function is used extensively in virtually every discipline where statistics are used, including medicine. The function is

$$f(x) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{x^2}{2}\right\}, \ -\infty < x < +\infty$$

and the graph is shown below:

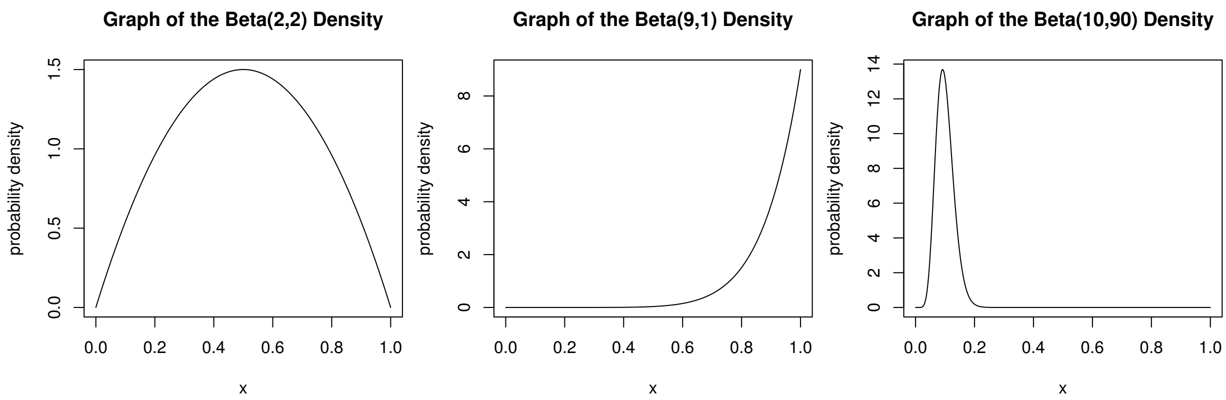


(iii) Another very common density used in Bayesian analysis is the beta. As we will see later in the course, it is typically used in problems involving proportions. Note that its range is between 0 and 1, very convenient for proportions. The function for the beta density is

$$f(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}, & 0 \leq \theta \leq 1, \ \alpha, \beta > 0, \ \text{ and} \\ 0, & \text{otherwise,} \end{cases} .$$

[$B(\alpha, \beta)$ represents the Beta function evaluated at (α, β). It is simply the normalizing constant that is necessary to make the density integrate to one, that is, $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$.] Some graphs of beta densities are shown below.

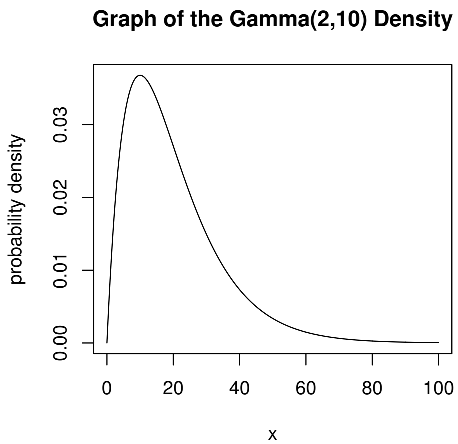
Note the flexibility of this family of distributions.



(iv) Yet another useful distribution is the gamma, which is sometimes used to model normal variances (or, more accurately, as we will see, the inverse of normal variances, known as the precision, i.e., precision = 1/variance). The gamma density is given by

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha - 1}, \ \ for\ x > 0\ .$$

A typical gamma graph is:



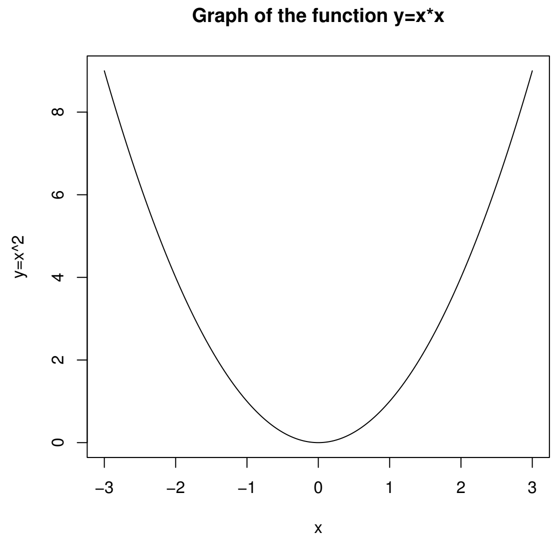
Derivatives: The *derivative* of a function measures the slope of the tangent line to the graph of the function at a given point. For example, if

$$f(x) = x^2,$$

then the derivative is given by

$$f'(x) = 2 \times x.$$

For example, this means that the slope of the tangent line at the point $x = 2$ (with $f(x) = y = 4$) is $2 \times 2 = 4$.



You may recall the following useful facts relating to derivatives:

1. The slope of a line is a measure of how quickly the function is rising or falling as x increases in value.

2. If a function has a maximum or minimum value, the the derivative is usually equal to 0 at that point. In the above, the function has a minimum at $x = 0$, where the value of the derivative is zero.

Derivatives are used in statistics for deriving maximum likelihood estimators, not used much in Bayesian analysis (at least not in this course). But the next topic is very important.

Integrals: The *indefinite integral* is a synonym for "anti-differentiation". In other words, when we calculate the indefinite integral of a function, we look for a function that when differentiated, returns the function under the integral sign. For example, the indefinite integral of the function $f(x) = x^2$ is given by the

$$\int x^2 \, dx = \frac{1}{3} \times x^3$$

because the derivative of $\frac{1}{3} \times x^3$ is x^2.

Indefinite integrals are used in many places in statistics, but as we will soon see, we use indefinite integrals to go from a *joint density* (many variables at once) to a *marginal density* (of a single variable, or some proper subset of the full set of variables).

Definite Integrals: The *definite integral* of a function is the area under the graph of that function. This area can be approximated directly from the graph, but exact mathematical formulae are also available from calculus. For example, the area under the the curve ranging from -1 to +2 of the function $f(x) = x^2$ is given by the following definite integral formula:

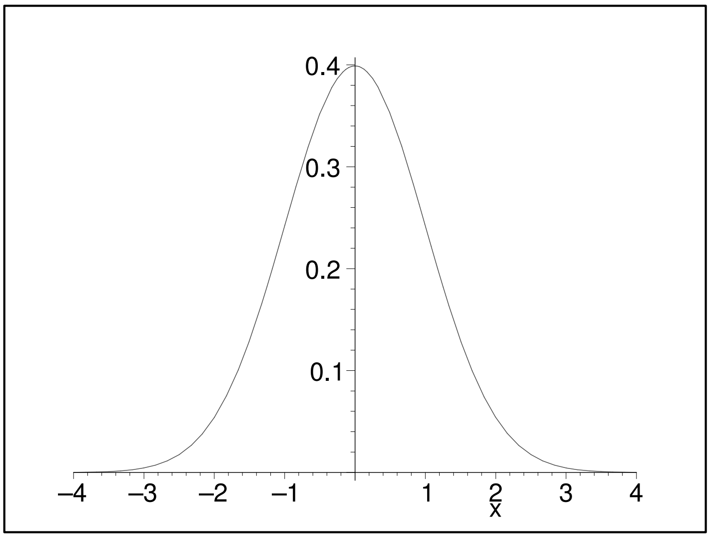
$$\int_{-1}^{+2} x^2 \, dx \;=\; \frac{1}{3} \times x^3 \Big|_{-1}^{+2} \;=\; \frac{2^3}{3} - \frac{(-1)^3}{3} \;=\; \frac{8}{3} + \frac{1}{3} \;=\; 3.$$

The area under a curve of a probability density function gives the probability of getting values in the region of the definite integral. For example, supposed we wished to calculate the probability that in choosing a random number between 0 and 1 (Uniform density function) the particular number we choose falls between 0.2 and 0.4. This is calculated by the definite integral

$$\int_{0.2}^{0.4} 1 \, dx \;=\; x \Big|_{0.2}^{0.4} \;=\; 0.4 - 0.2 \;=\; 0.2.$$

Definite integrals are also used in the context of calculating means and variances of random variables.

Joint and Marginal Distributions When we have only one parameter, we speak of its density. For example, if $x \sim N(0, 1)$, then the graph of the probability density is:



When we have two or more parameters, we speak of a *joint probability density*. For example, let x and y be *jointly multivariately* normally distributed, which is notated by:

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \Sigma \right)$$

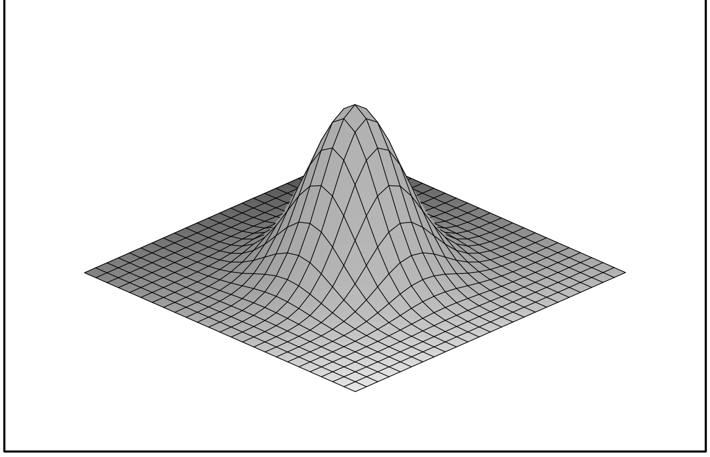
where

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \rho_{xy} \\ \rho_{xy} & \sigma_y^2 \end{pmatrix}$$

<u>Example:</u> Suppose

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$

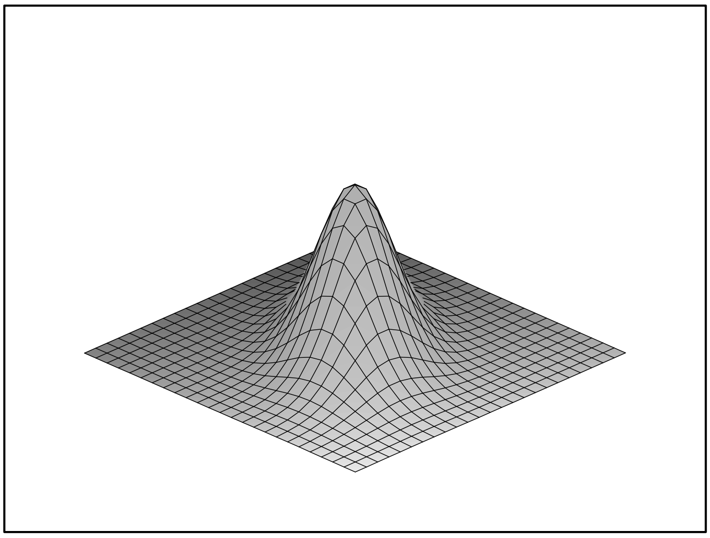
which is equivalent to two independently normally distributed variables, with no correlation between them. Then the picture is:



Note how the "slices" resemble univariate normal densities in all directions. These "slices" are marginal densities, which we will define later. In the presence of correlations, for example a correlation of 0.5, we have

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \right)$$

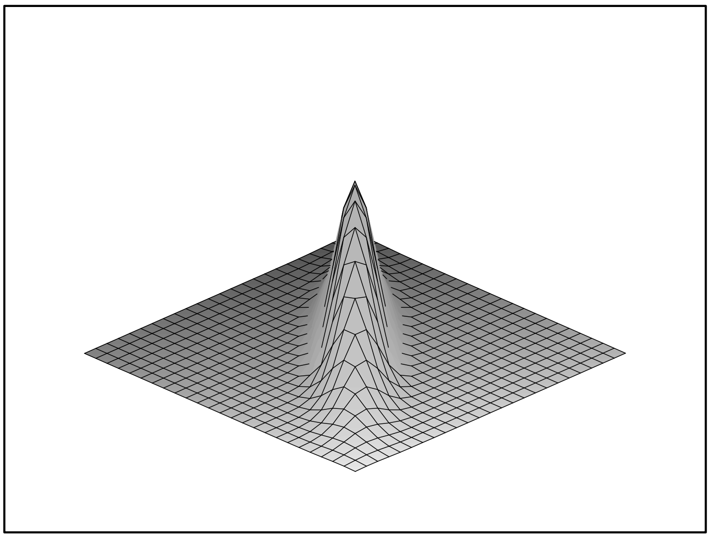
and the picture is:



Similarly, with very high correlation of 0.9, we have

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}\right)$$

and the picture is:



The *bivariate* normal density formula is:

$$f(x,y) = \frac{\exp\left\{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right\}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}$$

This is a joint density between two variables, since we look at the distribution of x and y at the same time, i.e., jointly. An example where such a distribution might be useful would be looking at both age and height together.

When one starts with a joint density, it is often of interest to calculate *marginal* densities from the joint densities. Marginal densities look at each variable one at a time, and can be directly calculated from joint densities through integration:

$$f(x) = \int f(x,y)dy, \text{ and}$$

$$f(y) = \int f(x,y)dx.$$

In higher dimensions,

$$f(x) = \int f(x,y,z)dydz,$$

and so on.

Normal marginals are normal If $f(x,y)$ is a bivariate normal density, for example, it can be proven that both the marginal densities for x and y are also normally distributed. For example, if

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{pmatrix} \sigma_x^2 & \rho_{xy} \\ \rho_{xy} & \sigma_y^2 \end{pmatrix}\right)$$

then

$$x \sim N(\mu_x, \sigma_x^2)$$

So, marginals from a multivariate normal distribution are always also normal.

Conditional Distributions Many of you have probably seen conditional densities defined for discrete variables, using definitions such as:

The conditional probability of event E given that event F has happened, is defined to be

$$P(E|F) = \frac{P(E \text{ and } F)}{P(F)}.$$

This is interpreted as "Given that F has occurred, calculate the probability E will also occur." Note that we can also write

$$P(E \text{ and } F) = P(F) \times P(E|F),$$

even if E and F are not independent.

There is a similar rule for continuous densities, which can be stated as:

The conditional *density* of random variable x given the value of a second random variable y is defined to be:

$$f(x|y) = \frac{f(x, y)}{f(y)}$$

Note the similarities between the discrete and continuous cases. If you have three or more variables, similar definitions apply, such as:

$$\begin{aligned} f(x|y, z) &= \frac{f(x, y, z)}{f(y, z)} \\ f(x, y|z) &= \frac{f(x, y|z)}{f(z)} \end{aligned}$$

and so on. The concept of conditional distributions is very important to modern Bayesian analysis since they are key in algorithms such as the Gibbs sampler.

Summary:

- Joint densities describe multi-dimensional probability distributions for two or more variables.
- If one has a joint density, then if it is of interest to look at each variable separately, one can find marginal probability distributions by integrating the joint densities. If one wants the marginal distribution of x, for example, then one would "integrate out" all of the parameters except x, and so on.
- For multivariate normal distributions, all marginal densities are again normal distributions, with the same means and variances as the variables have in the joint density.
- The concept of conditionality applies to continuous variables.