McGill University

Department of Epidemiology and Biostatistics

Bayesian Analysis for the Health Sciences

Course EPIB-675

Lawrence Joseph

Bayesian Analysis for the Health Sciences – EPIB-675 – 3 credits

Instructor:	Lawrence Joseph
Email address:	Lawrence.Joseph@mcgill.ca (best way to reach me)
Home page:	http://www.medicine.mcgill.ca/epidemiology/Joseph/
Telephone:	934-1934 X 44713
Address:	Division of Clinical Epidemiology Royal Victoria Hospital V Building Room V2.10

Course Objectives and Topics Covered: To provide researchers with an introduction to practical Bayesian methods. Topics will include Bayesian philosophy, simple and more complex models, linear and logistic regression, hierarchical models, diagnostic tests, sample size methods, issues in clinical trials, measurement error and missing data problems. Numerical techniques including Monte Carlo integration, sampling importance resampling (SIR), and the Gibbs sampler will be covered, including programming in R and WinBUGS. While all examples will be to epidemiological research, most of the ideas and material will be applicable to other areas of research.

Place and Time: January 9 to April 16, 2012. Mondays and Wednesdays, 11:00 AM to 12:30 PM. Room 25, Purvis Hall, 1020 Pine Avenue West, corner Peel Street.

Assessment: Ten assignments of approximately 5 questions each, about one per week throughout the term. Each assignment is worth 10%. There will be no exams.

Textbook: A. Gelman, J. Carlin, H. Stern and D. Rubin, Bayesian Data Analysis, 2nd Edition, Chapman and Hall, 2003.

Prerequisites: At least two previous courses in statistics, including topics such as inferences for means and proportions, and linear and logistic regression. Differential and integral calculus. If you are unsure you have sufficient background, please speak to the instructor.

Bayesian Analysis in the Health Sciences

Course Outline – EPIB–675, January – April 2012

Date	Topic Covered
Mon Jan 9	Introduction/Motivation/Evaluation/Scope
Wed Jan 11	Multivariate Distributions, Conditionality
Mon Jan 16	Basic Elements of Bayesian Analysis
Wed Jan 18	Bayesian Philosophy I
Mon Jan 23	Bayesian Philosophy II
Wed Jan 25	Simple Models I - Univariate Models
Mon Jan 30	Simple Models II - Predictive Distributions
Wed Feb 1	Computation and Numerical Methods I - Introduction
Mon Feb 6	Computation and Numerical Methods II - Monte Carlo Integration
Wed Feb 8	Computation and Numerical Methods III - SIR Algorithm
Mon Feb 13	Computation and Numerical Methods IV - Gibbs sampler and WinBUGS
Wed Feb 15	Computation and Numerical Methods V - More on WinBUGS
Mon Feb 20	No Class – Spring Break
Wed Feb 22	No Class – Spring Break
Mon Feb 27	Bayesian Linear and Logistic Regression
Wed Feb 29	Hierarchical Linear and Logistic Regression
Mon Mar 5	Bayesian Analysis of Clinical Trials
Wed Mar 7	Hierarchical Models I - Simple Hierarchical Models
Mon Mar 12	Hierarchical Models II - Meta Analysis with Random Effects
Wed Mar 14	Hierarchical Models III - More Complex Hierarchical Models
Mon Mar 19	Adjusting for Measurement Error
Wed Mar 21	Prior Distributions - Prior Selection and Elicitation
Mon Mar 26	Model Selection in Regression - Bayes Factors
Wed Mar 28	Missing Data
Mon Apr 2	Bayesian bias adjustments
Wed Apr 4	Bayesian Sample Size Criteria
Mon Apr 9	No Class – Easter Monday
Wed Apr 11	Analysis of Diagnostic Test Data
Mon Apr 16	Discussion and Conclusions - The Future of Bayesian Analysis

Bayesian Probabilities - Discrete Case of Bayes Theorem

It is easy to get confused between Bayesian analysis as an inferential paradigm, and Bayes Theorem as a basic way to manipulate discrete probabilities. Let us first consider the discrete case:

Suppose we are considering a test for cancer:

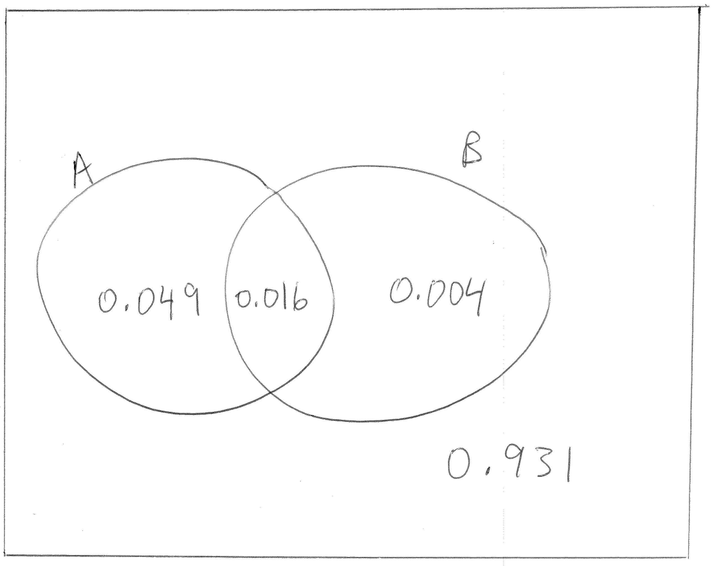
Let A = the event that a test is positive.
Let B = the event of actually having cancer.

Suppose we know that:

- $P(A|B^c) = 0.05$, and so $P(A^c|B^c) = 1 - 0.05 = 0.95$
- $P(A^c|B) = 0.20$, and so $P(A|B) = 1 - 0.20 = 0.80$
- $P(B) = 0.02$, and so $P(B^c) = 0.98$

(a) What is the probability of cancer given that the test is positive?
(b) What is the probability of cancer given that the test is negative?

We can draw a diagram as below:



From the diagram, we see that

$$P(B|A) = \frac{0.016}{0.016 + 0.049} = .2462$$

and

$$P(B|A^c) = \frac{0.004}{0.004 + 0.931} = .0043$$

Alternatively, we can use Bayes Theorem, which states:

$$P(B|A) = \frac{P(B) \times P(A|B)}{P(B) \times P(A|B) + P(B^c) \times P(A|B^c)}$$

Plugging in the numbers, we can check that the solutions are the same. For example,

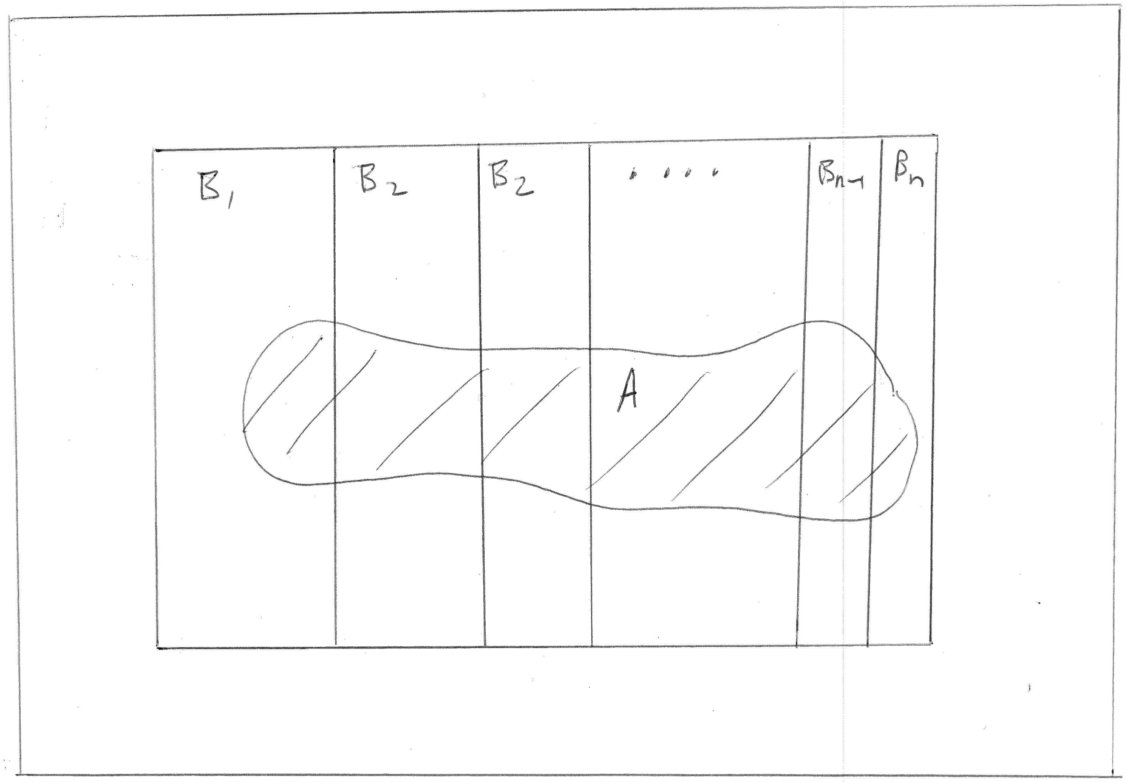
$$P(B|A) = \frac{P(B) \times P(A|B)}{P(B) \times P(A|B) + P(B^c) \times P(A|B^c)} = \frac{0.02 \times 0.80}{0.02 \times 0.80 + 0.98 \times 0.05} = .2462.$$

Switching the roles of A and A^c in the above formula yields

$$P(B|A^c) = \frac{P(B) \times P(A^c|B)}{P(B) \times P(A^c|B) + P(B^c) \times P(A^c|B^c)} = 0.0043$$

Note that before the test is performed, the probability that a person has cancer is 0.02, but that these probabilities are "updated" in a natural way, once the test results become available.

Bayes Theorem may be generalized to the case where the event B has more than two possible outcomes, say $B_1,\ B_2, \ldots, B_n$.



In this case, the Bayes Theorem is

$$P(B_k|A) = \frac{P(B_k) \times P(A|B_i)}{\sum_{i=1}^{n} P(B_i) \times P(A|B_i)}, \ k = 1, 2, \ldots, n.$$

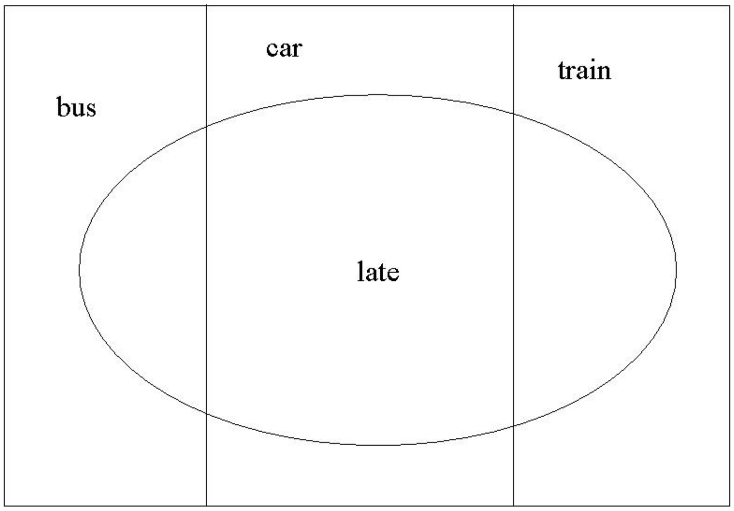
Here is an example for this case:

Suppose that Bob can decide to go to work by one of three modes of transportation, car, bus, or commuter train. Because of high traffic, if he decides to go by car, there is a 50% chance he will be late. If he goes by bus, which has special reserved lanes but is sometimes overcrowded, the probability of being late is only 20%. The commuter train is almost never late, with a probability of only 1%, but is more expensive than the bus.

(a) Suppose that Bob is late one day, and his boss wishes to estimate the probability that he drove to work that day by car. Since he does not know which mode of transportation Bob usually uses, he gives a prior probability of $\frac{1}{3}$ to each of the three possibilities. What is the boss' estimate of the probability that Bob drove to work?

(b) Suppose that a coworker of Bob's knows that he almost always takes the commuter train to work, never takes the bus, but sometimes, 10% of the time, takes the car. What is the coworkers probability that Bob drove to work that day, given that he was late?

Solution: The Venn diagram would be:



(a) We have the following information given in the problem:

$$\begin{aligned} Pr\{ \text{ bus } \} = Pr\{ \text{ car } \} = Pr\{ \text{ train } \} &= \frac{1}{3} \\ Pr\{ \text{ late } | \text{ car } \} &= 0.5 \\ Pr\{ \text{ late } | \text{ train } \} &= 0.01 \\ Pr\{ \text{ late } | \text{ bus } \} &= 0.2 \end{aligned}$$

We want to calculate $Pr\{$ car | late $\}$.

By Bayes Theorem, this is

$$
\begin{aligned}
& Pr\{ \text{ car } | \text{ late } \} \\
= & \frac{Pr\{ \text{ late } | \text{ car } \} Pr\{ \text{ car } \}}{Pr\{ \text{ late } | \text{ car } \} Pr\{ \text{ car } \} + Pr\{ \text{ late } | \text{ bus } \} Pr\{ \text{ bus } \} + Pr\{ \text{ late } | \text{ train } \} Pr\{ \text{ train } \}} \\
= & \frac{0.5 \times 1/3}{0.5 \times 1/3 + 0.2 \times 1/3 + 0.01 \times 1/3} \\
= & 0.7042
\end{aligned}
$$

(b) Repeat the identical calculations as the above, but instead of the prior probabilities being $\frac{1}{3}$, we use $Pr\{$ bus$\} = 0$, $Pr\{$car$\} = 0.1$, and $Pr\{$ train $\} = 0.9$. Plugging in to the same equation with these three changes, we get $Pr\{$ car | late $\} = 0.8475$

This is a simple theorem in probability, having nothing to do with drawing inferences from a data set, that *everybody* uses. Bayes Theorem creates no controversy whatsoever (not that Bayesian inference is so controversial nowadays).

Bayesian Inference - Continuous Case of Bayes Theorem

The above discrete version is different from the continuous version of Bayes Theorem, in that it is typically used for drawing inferences, as an alternative to the freqeuntist approach that leads to p-values and confidence intervals. The continuous version of Bayes Theorem looks like this:

$$\text{posterior distribution} = \frac{\text{likelihood of the data} \times \text{prior distribution}}{\text{a normalizing constant}},$$

or

$$f(\theta|x) = \frac{f(x|\theta) \times f(\theta)}{\int f(x|\theta) \times f(\theta) d\theta,}$$

or, forgetting about the normalizing constant,

$$f(\theta|x) \propto f(x|\theta) \times f(\theta).$$

Thus we "update" the prior distribution to a posterior distribution after seeing the data via Bayes Theorem.

We will see many examples of its use later in the course.

Aspirin		Tylenol	
Cured	Not Cured	Cured	Not Cured
5	5	5	5
6	4	5	5
6	4	4	6
7	3	4	6
8	2	4	6
8	2	3	7
9	1	3	7
⋮	⋮	⋮	⋮
10	0	0	10

MOZART
BEETOVEN
TEA
MILK
1?
2?
HEADS!
HEADS!
HEADS
TAILS!

Effects of Therapeutic Touch On Tension Headache Pain

ELIZABETH KELLER • VIRGINIA M. BZDEK

Therapeutic touch (TT) is a modern derivative of the laying on of hands that involves touching with the intent to help or heal. This study investigated the effects of TT on tension headache pain in comparison with a placebo simulation of TT. Sixty volunteer ~~subjects with ten~~sion headaches were randomly divided into treatment and placebo groups. The McGill-Melzack Pain Questionnaire was ~~used to measure headache~~ pain levels before each intervention, immediately afterward, and 4 hours later. A Wilcoxon signed rank test for differences indicated that 90% of the subjects exposed to TT experienced a sustained reduction in headache pain, p < *.0001. An average 70% pain reduction was sustained over the 4 hours following TT, which was twice the average pain reduction following the placebo touch. Using a Wilcoxon rank sum test, this was statistically significant,* p < *.01. Study results indicated that TT may have potential beyond a placebo effect in the treatment of tension headache pain.*

Therapeutic touch (TT), a modern version of the laying on of hands, was introduced into nursing by Krieger (1975). It does not entail belief in the method or in any other precept on the part of its recipients to be effective (Krieger, 1979). TT may or may not involve contact with the physical body, but contact is said always to be made with the energy field of the healing process (Boguslawski, 1979; Krieger, 1975, 1981).

Background of the Study

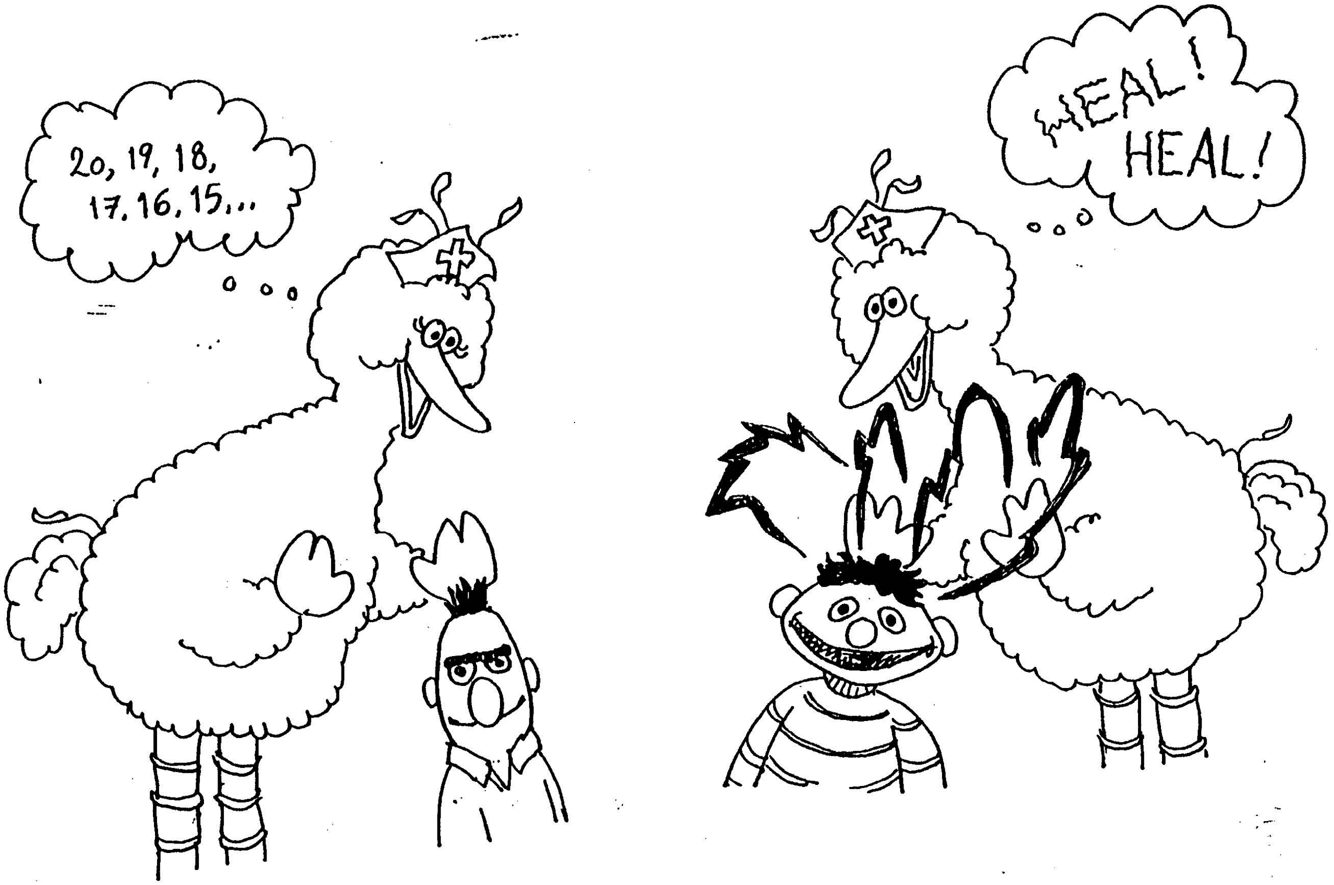
Therapeutic touch is based on the philosophy of holism (Krieger, 1981; Weber, 1981) and general systems theory (Battista, 1977). Holism is represented in nursing science by Roger's (1970) theory of unitary man. According to this theory, all persons are highly complex fields of various forms of life energy. These fields of energy are coextensive with the universe and in constant interaction and exchange with surrounding energy fields. The functional basis of TT lies in the direction of life energy through the hands of the therapist to the recipient who may then internalize this energy, use it to restore balance, and thereby self-heal (Boguslawski, 1979; Krieger, 1979, 1981). The predominant theory in recent TT literature concerning the source of the transferred energy is that the therapist serves as a conduit, a channel, so that environmental energy may be transferred to the recipient (Boguslawski, 1979; Weber, 1981). To be recognized as a realistic and tenable phenomenon TT must be considered within a holistic context.

Nurse researchers have investigated TT since Krieger (1976) demonstrated increased hemoglobin

of three groups of 30 hosp
diovascular patients: the
TT, casual touch, or nc
group receiving TT show
cant reduction in state a
on the Spielberger Se
Questionnaire (Spielberg
& Lushene, 1970) p
compared with their
scores and with the post
the other two groups, *p*

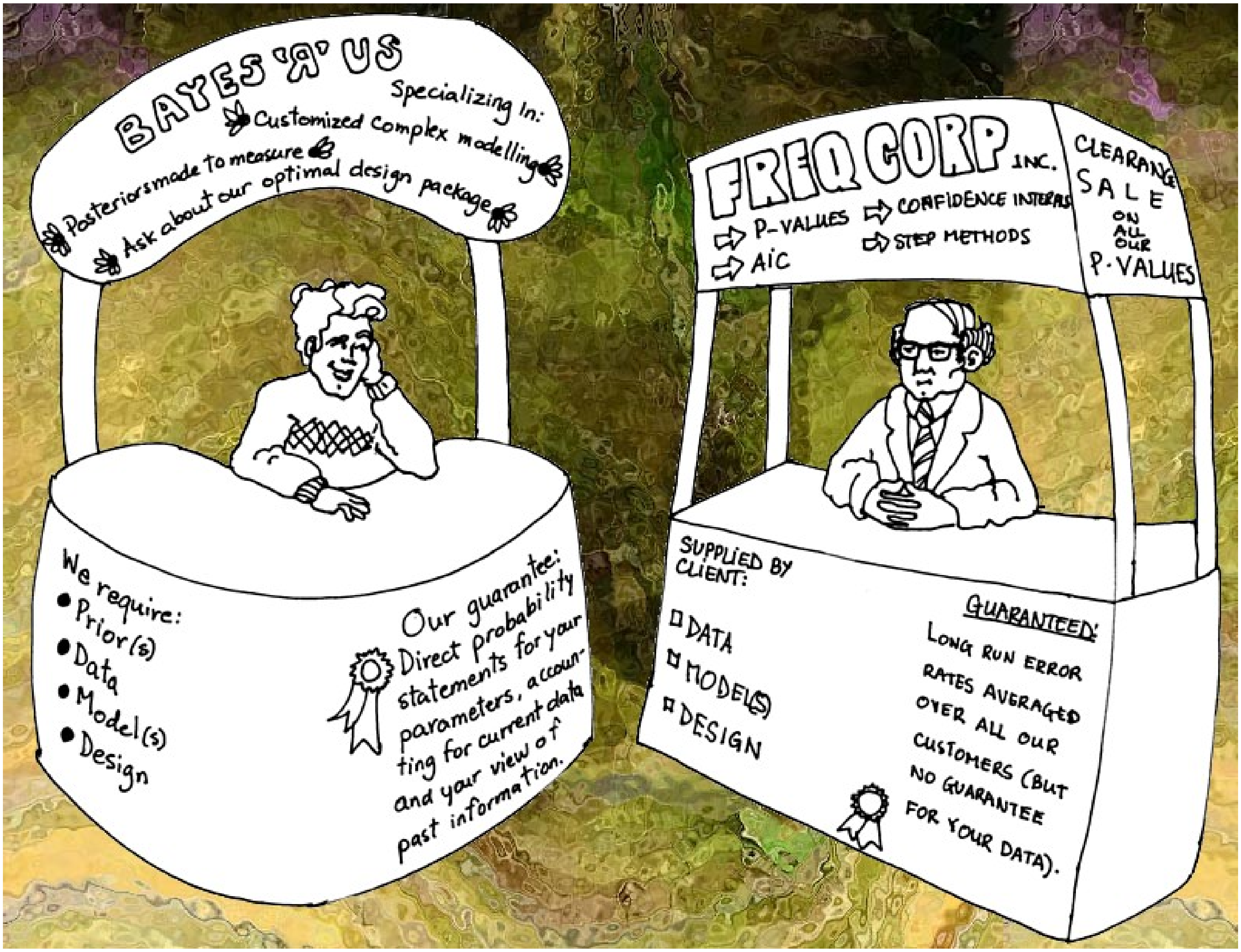
Quinn (1982) replic
anxiety study with 60 c
patients, but replaced
(pulse-taking) with pl
which was a simulation
ments without energy t
of the subjects received
tact during either
Quinn reasoned that if
for TT is an exchange
between human fields
beyond the skin, direct
tact would not be nec
indicated that the n
group demonstrated
lower posttest anxiety s
noncontact placebo g
degree of anxiety
Quinn's noncontact T
almost identical to He
the physical contact T

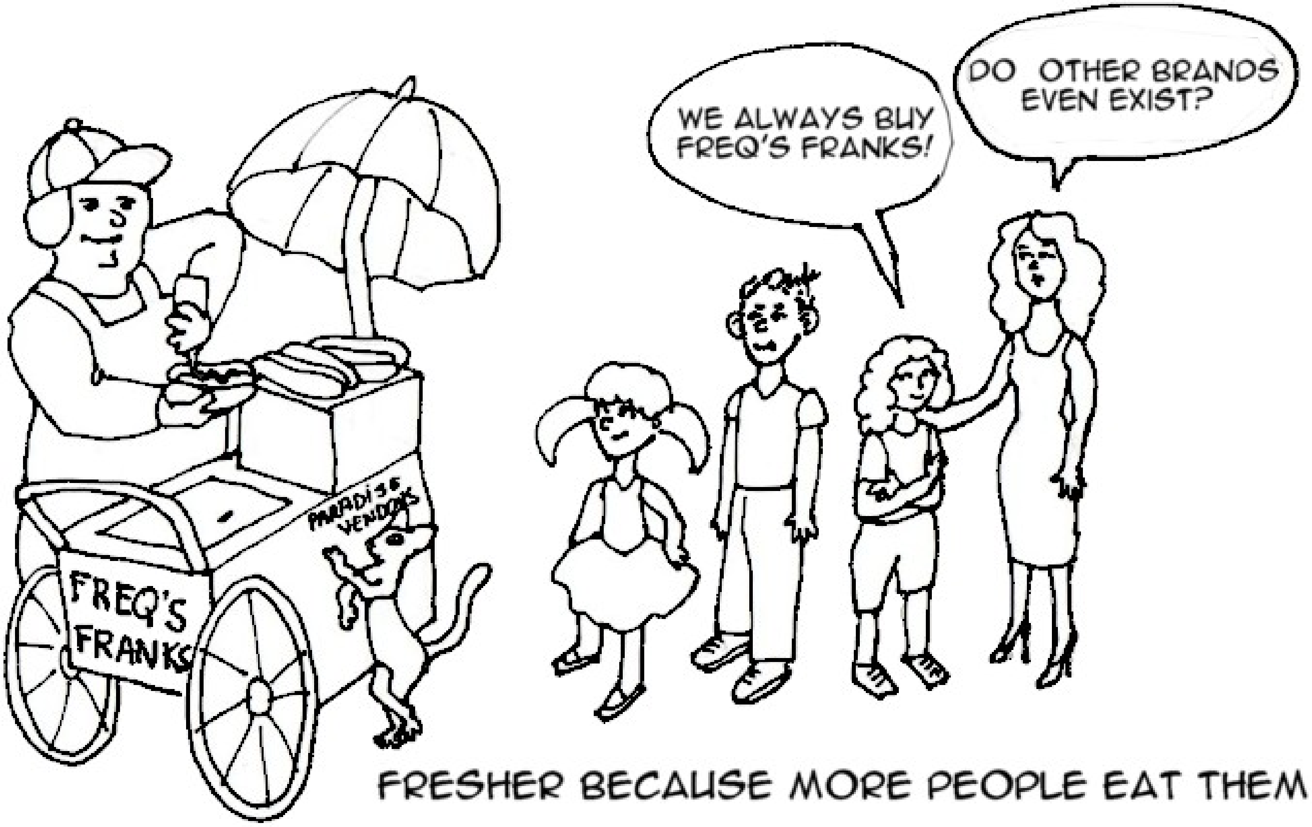
Randolph (1984)
physiologic response
college students to a st
film while receiving ei


20, 19, 18,
17. 16, 15...
HEAL!
HEAL!

Neigh!

Definitely, it's a cocktail!
Nah, it's seafood.
No way. It's a source of nitrites.


BAYES 'R' US
Specializing In:
Customized complex modelling
Posteriors made to measure
Ask about our optimal design package
We require:
● Prior(s)
● Data
● Model(s)
● Design
Our guarantee: Direct probability statements for your parameters, accounting for current data and your view of past information.
FREQ CORP INC.
⇨ P-VALUES
⇨ AIC
⇨ CONFIDENCE INTERVALS
⇨ STEP METHODS
CLEARANCE SALE ON ALL OUR P-VALUES
SUPPLIED BY CLIENT:
□ DATA
□ MODEL(S)
□ DESIGN
GUARANTEED:
LONG RUN ERROR RATES AVERAGED OVER ALL OUR CUSTOMERS (BUT NO GUARANTEE FOR YOUR DATA).



FRESHER BECAUSE MORE PEOPLE EAT THEM

MORE PEOPLE EAT THEM BECAUSE THEY ARE FRESHER

Mathematical Background

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

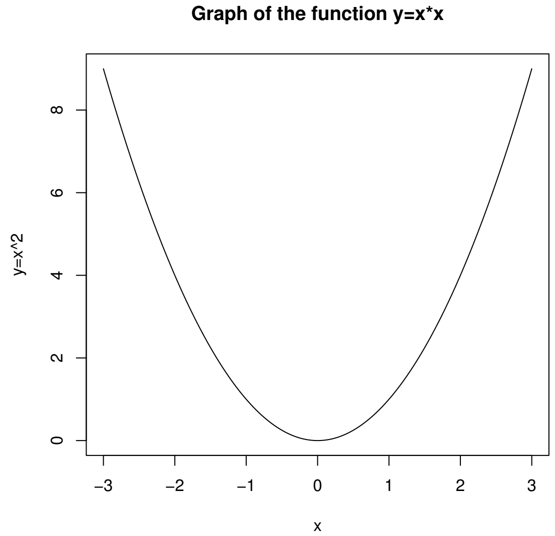
Note: The following are *very* non-rigorous definitions designed to suit the purpose of our course. Refer to any calculus and/or statistics textbook for the exact definitions and/or more information.

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.

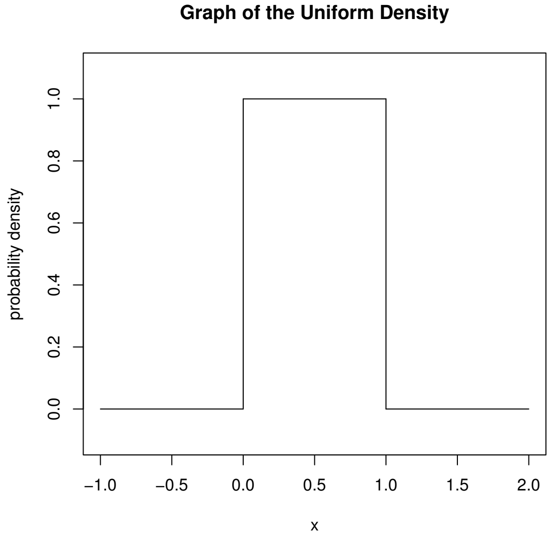


Functions are used in statistics to describe probability (density) functions (among many other things). Some examples:

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

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

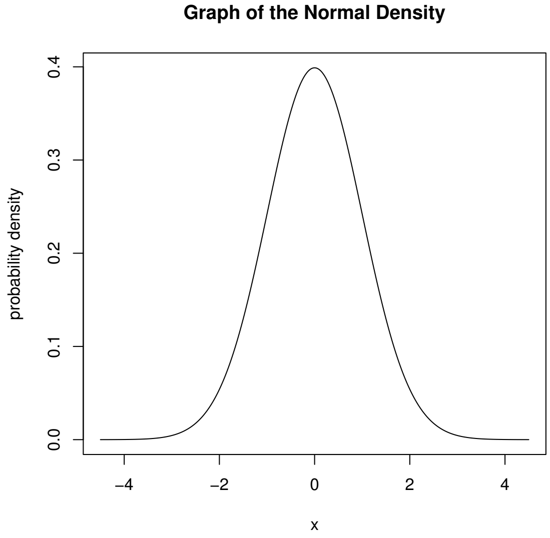
and the graph is shown below:



(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:

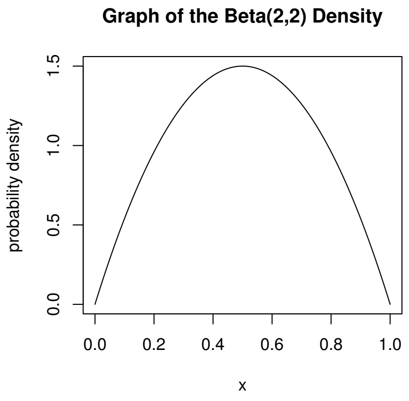


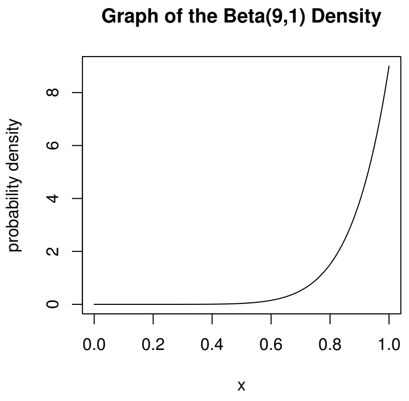
(iii) Another very common density used in Bayesian analysis is the beta. As we will see later in the course, it is typically used in problems involving proportions. Note that its range is between 0 and 1, very convenient for proportions. The function for the beta density is

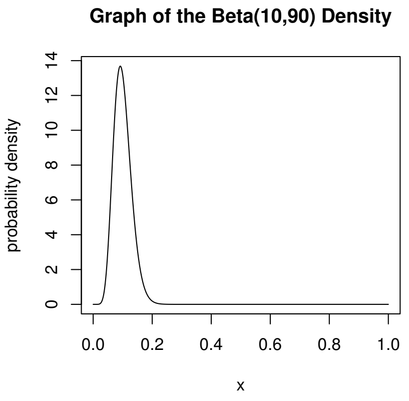
$$f(x) = \left\{ \begin{array}{ll} \frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}, & 0 \leq \theta \leq 1, \ \alpha, \beta > 0, \ \text{ and} \\ 0, & \text{otherwise,} \end{array} \right. .$$

[$B(\alpha, \beta)$ 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.


Graph of the Beta(2,2) Density

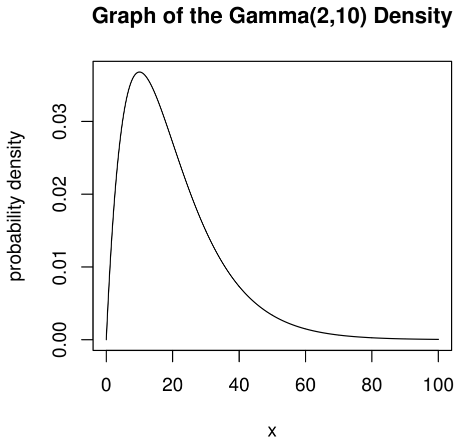

Graph of the Beta(9,1) Density


Graph of the Beta(10,90) Density

(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:


Graph of the Gamma(2,10) Density

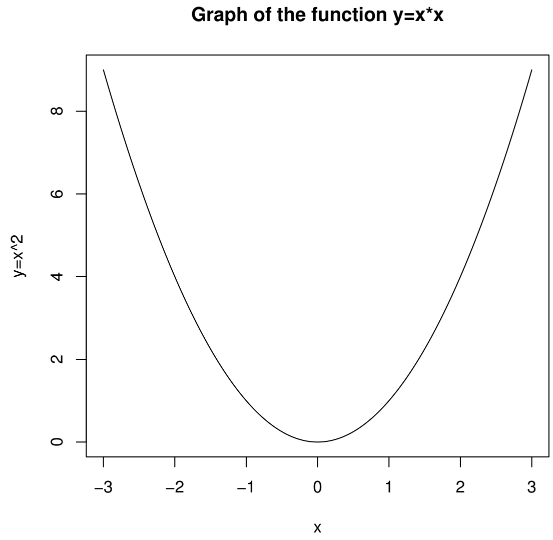
Derivatives: The *derivative* of a function measures the slope of the tangent line to the graph of the function at a given point. For example, if

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

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:

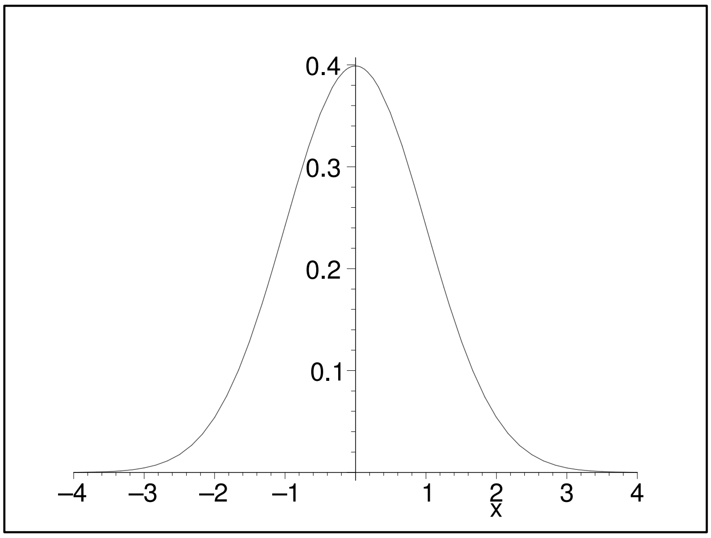
$$\int_{-1}^{+2} x^2 \, dx \;=\; \frac{1}{3} \times x^3 \Big|_{-1}^{+2} \;=\; \frac{2^3}{3} - \frac{(-1)^3}{3} \;=\; \frac{8}{3} + \frac{1}{3} \;=\; 3.$$

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:

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

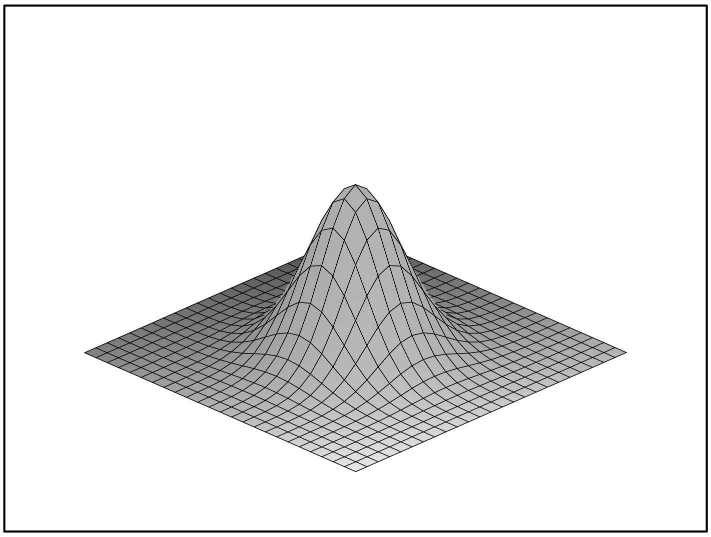
where

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

<u>**Example:**</u> Suppose

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

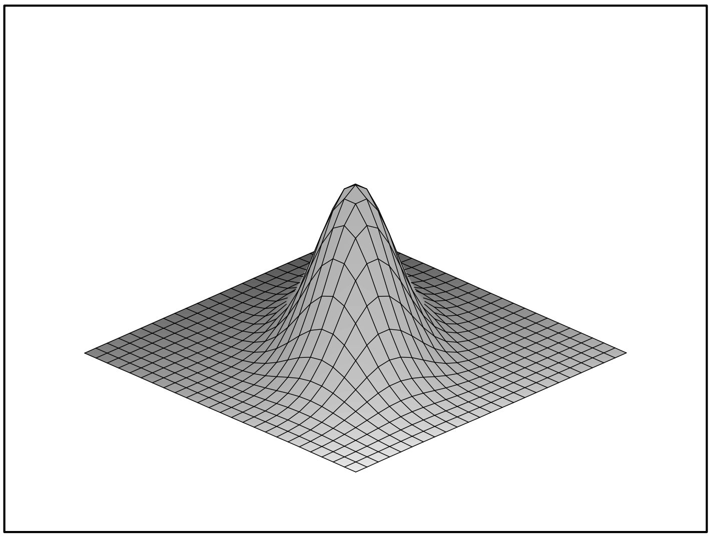
which is equivalent to two independently normally distributed variables, with no correlation between them. Then the picture is:



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

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

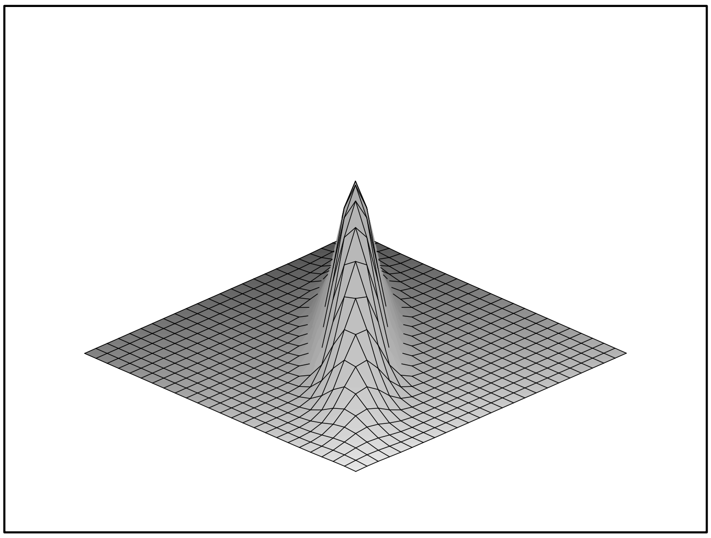
and the picture is:



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.