9.07 Introduction to Statistics for Brain and Cognitive Neuroscience
Emery N. Brown

Lecture 3 Continuous Probability Models

I. Objectives

1. **Understand the concept of a continuous probability model and a continuous random variable.**

2. **Understand the Inverse Transformation Method for simulating a random variable.**

3. **Understand the five-number summary and boxplots as Exploratory Data Analysis methods of summarizing a sample of data.**

4. **Understand the central role of the Gaussian distribution in probability modeling and statistical inference.**

5. **Become familiar with Q-Q plots as a way of assessing goodness-of-fit.**

6. **Become familiar with other continuous probability models that are relevant for statistical analyses in neuroscience.**

    a. **Exponential**
    b. **Gamma**
    c. **Beta**
    d. **Inverse Gaussian**

II. Continuous Probability Models

In **Lecture 2**, we studied discrete probability models. In this lecture we study a set of continuous probability models. The principal probability model we will study in this lecture is the Gaussian.

A. Continuous Probability Distributions

For a continuous probability model we have from our **Axioms of Probability**

$$\int_A^B f(x)dx = 1. \tag{3.1}$$

Any positive function $g(x)$ that has a finite integral on an interval (A,B) may be converted into a probability density by computing

$$\int_A^B g(x)dx = C \tag{3.2}$$

and defining simply

$$f(x) = \frac{g(x)}{C}. \tag{3.3}$$

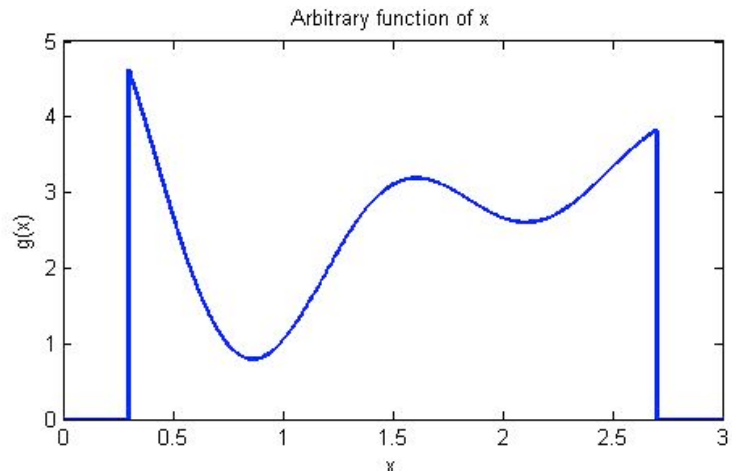


Figure 3A. A positive valued-function g(x).

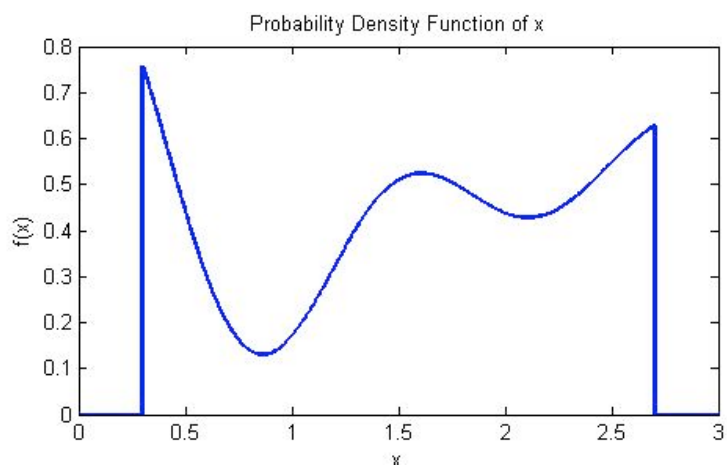


Figure 3B. Probability density function constructed from g(x) in Figure 3A.

We term $f(x)$ the **probability density function** (pdf).

B. Continuous Random Variables (Kass, 2006).

A **random variable** is **continuous** if it takes on all values in some interval (A,B), where it is possible that either $A=-\infty$ or $B=\infty$ (or both). The mathematical distinction between discrete and continuous distributions is that: (i) discrete distributions take on only certain specific values (such as non-negative integers) that can be separated from each other (thus the name discrete); and (ii) wherever summation signs appear for discrete distributions, integrals replace them for continuous distributions. If X is a continuous random variable its **probability density function** (pdf) will be written as $f(x)$ where now

$$\Pr(a \le X \le b) = \int_a^b f(x)dx \tag{3.4}$$

and as mentioned above, the integral over the entire domain of $f(x)$ is 1. Note that in this continuous case there is no distinction between $\Pr(a \le X)$ and $P(a < X)$ because $\Pr(X = a) = 0$. We may think of $f(x)$ as the probability per unit of X and hence, $f(x)dx = P(x \le X \le x + dx)$. *N.B.* $\Pr(x = a) = 0$ does not imply that a is impossible. Why?

Example. 3.1. Uniform Probability Density. Perhaps the simplest example is the uniform distribution. For instance, if the time of day at which births occurred followed a uniform distribution, then the probability of a birth in any given 30 minute period would be the same as that for any other 30 minute period throughout the day. In this case, the pdf $f(x)$ would be constant over the interval from 0 to 24 hours. Because it must integrate to 1, we must have $f(x) = 1/24$ and the probability of a birth in any given 30 minute interval starting at a hours is

$$\int_a^{a+0.5} f(x)dx = 1/48. \tag{3.5}$$

The **cumulative distribution function** (cdf) (or simply *distribution function*) is written again as $F(x)$ and is defined as before: $F(x) = P(X \le x)$. This becomes

$$F(x) = \int_A^x f(u)du. \tag{3.6}$$

Also, note that by the Fundamental Theorem of Calculus $f(x) = F'(x)$.

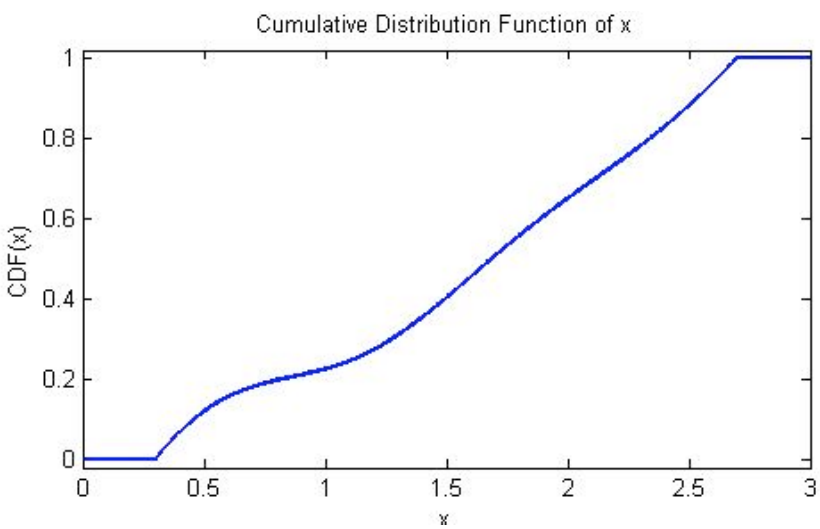


Figure 3C. Cumulative Distribution Function for $f(x)$ in Figure 3B.

The **quantiles** or **percentiles** are often used in working with continuous distributions: for p a number between 0 and 1 (such as 0.25), the pth quantile or $100p-th$ percentile (e.g. the 0.25 quantile or the 25th percentile) of a distribution having cdf $F(x)$ is the value η such that

$p = F(\eta)$. The median is the Thus, we may write $\eta = F^{-1}(p)$, where F^{-1} is the inverse **cumulative distribution function** or inverse cdf.

C. Simulating a Random Variable

Most statistical software contains routines to generate artificially random observations from standard distributions, such as those presented in these notes. These are really *pseudo*-random number generators because they are created by algorithms that are actually deterministic, so that in very long sequences they repeat and their non-random nature becomes apparent. A cdf F outputs a number between 0 and 1. Therefore if F is continuous and if we take U as a uniform random variable on $(0,1)$, then we can define

$$X = F^{-1}(U) \tag{3.7}$$

Where $F^{-1}(u)$ equals x whenever $F(x) = u$. Because $F(x)$ is continuous and monotonic, it follows that we can simulate a random variable X from the continuous cdf F, whenever F^{-1} is computable by simulating a random number U, then setting $X = F^{-1}(U)$. These observations imply the following algorithm known as the **Inverse Transformation Method**.

Algorithm 3.1 (Inverse Transformation Method)

1. Pick n large. For $i = 1$ to n.
2. Draw U_i from $U(0,1)$ and compute $X_i = F^{-1}(U_i)$
3. The observations $(X_1, ..., X_n)$ are a random sample from F.

Example 3.2. Simulating Exponential Random Variables. We define below in **3H.1** the exponential probability density as

$$f(x) = \lambda e^{-\lambda x} \tag{3.8}$$

for $x > 0$ and $\lambda > 0$. It follows that the *cdf* $F(x)$ is

$$F(x) = \int_0^x \lambda e^{-\lambda u} du = 1 - e^{-\lambda x}. \tag{3.9}$$

To find $F^{-1}(u)$ we note that

$$1 - e^{-\lambda x} = u \tag{3.10}$$

$$x = -\frac{\log(1-u)}{\lambda} \tag{3.11}$$

Hence, if u is uniform on $(0,1)$ then

$$F^{-1}(u) = -\frac{\log(1-u)}{\lambda} \tag{3.12}$$

is exponentially distributed with mean λ^{-1}. Since $1-u$ is uniform on $(0,1)$, it follows that $x=-\frac{\log(u)}{\lambda}$ is exponential with mean λ^{-1}. Therefore, we can use **Algorithm 3.1** to simulate draws from an exponential distribution by using F^{-1} in Eq. 3.12.

D. Properties of a Random Variable
The **expected value** of X is

$$\mu_x = E(X) = \int_A^B xf(x)dx \tag{3.13}$$

and the **standard deviation** of X is $\sigma_x = \sqrt{Var(X)}$ where

$$\sigma_x^2 = Var(X) = \int_A^B (x-\mu_x)^2 f(x)dx \tag{3.14}$$

is the **variance** of X. In each of these formulas, we have simply replaced sums by integrals in the analogous definitions for discrete random variables in **Lecture 2**.

The following properties are often useful

$$\begin{aligned} &E(aX+b) = aE(X)+b \\ &Var(aX+b) = a^2 Var(X) \\ &\sigma_{ax+b} = |a|\sigma_x \end{aligned} \tag{3.15}$$

They are easy to prove. For example

$$\begin{aligned} E(aX+b) &= \int (ax+b)f(x)dx \\ &= a\int xf(x)dx + b\int f(x)dx \\ &= aE(X)+b \end{aligned} \tag{3.15}$$

If X is a continuous random variable with probability density function $f(x)$, then for any real-valued function g,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx. \tag{3.16}$$

We are most interested in

i) $g(x) = ax+b$
ii) $g(x) = (x-\mu)^2$ where $\mu = E[X]$
iii) $g(x) = x^n$ $n \geq 1$

From i) and the definition of the expectation it follows that

$$E(g(X)) = E[aX+b] = aE[X]+b.$$

We define the variance of a continuous random variable X as

$$Var(X) = E(X-\mu)^2 = E(X-E[X])^2.$$

The variance measures the expected squared deviation of the random variable from its mean. It provides a measure of spread of the probability mass function or probability density function. If $\mu = E[X]$ then it is easy to show that for continuous random variables that

$$Var(X) = E[X-\mu]^2 = E[X^2] - \mu^2.$$

D. An Exploratory Data Analysis Interlude

Example 3.3. Magnetoencephalography. Magnetoencephalography is a frequently used tool for studying brain function for clinical purposes and for cognitive studies. It is based on the idea that electrophysiological activity in the brain generates a small magnetic field from the spatial superposition of individual neuronal source currents. At a distance of about 15 mm from the scalp, the observed field is of the order of 10^{-13} to 10^{-12} Tesla peak-to-peak. This measurement process is termed magnetoencephalography (MEG). To minimize instrumental noise, the MEG is usually detected using superconducting flux transformers, coupled to superconducting quantum interference device (SQUID) sensors. Because the brain's magnetic field is small relative to those from the environment it must be measured in a specially designed, highly shielded room. Characterizing the background noise is an important step for calibrating the instrument. The characterization of the background noise is also used in solving the source localization problem.

On each channel of a magnetoencephalogram, we measure the magnetic field in an MEG scanner in the absence of a subject in order to characterize the background noise. An example of the data recorded during 14 minutes is shown in Figure 3D. The measurements were recorded at 601 Hz. What probability model would be appropriate for these data?

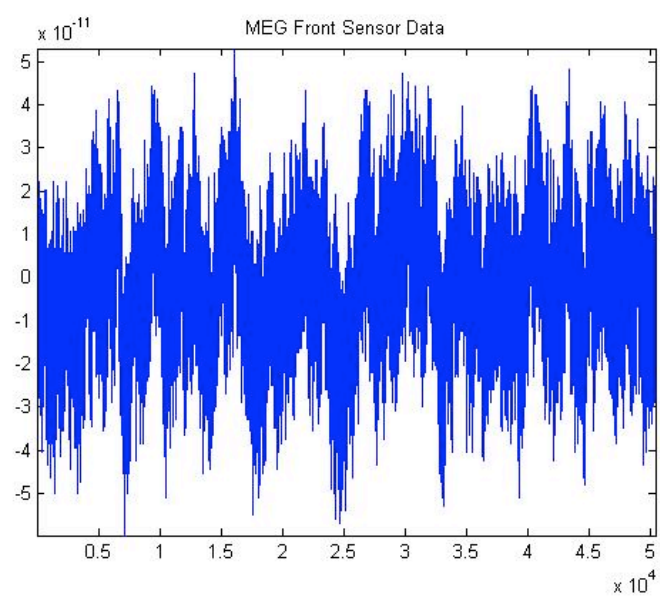
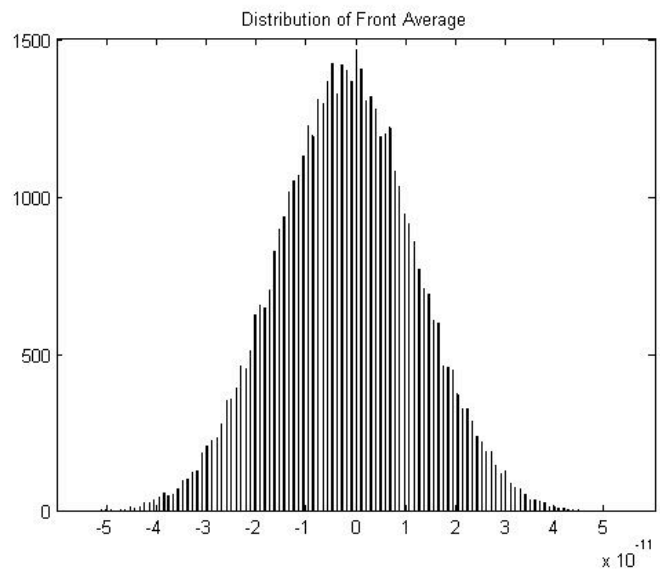


Figure 3D. Time-Series Plot of Magnetoencephalogram Background Noise Measurements (Courtesy of Simona Temereanca, MGH).

Before going on to propose a probability model for these data we plot them first as a histogram and then secondly as a boxplot. The histogram shows that the data are highly symmetric, bell-shaped and centered near 0 (Figure 3E). The **boxplot and whiskers plot** of the data is shown in Figure 3F along with the **five-number summary**. The boxplot is constructed by first finding the **median** or 50th percentile of the data. It is the middle value of all the observations. If there is an even number of observations, we take the median to be, by convention, the average of the two middle values. The lower border of the box is the 25th percentile and the upper border of the box is the 75th percentile. The median is placed as a horizontal line in the appropriate position in the box. That is, at the location in the box where it would fall. Note that this will not be in the middle of the box if the distribution is asymmetric. The distance between the 25th and 75th percentile is called the **interquartile range** (IQR). The **whiskers** extend from the upper and lower borders of the box and are taken to be either 1.5 or 3 times the IQR. Horizontal lines are placed at the end of each whisker to demarcate the primary extent of the box plot. Any data points which lie outside the whiskers are simply marked as horizontal lines and define outliers.



Tesla

Figure 3E. Histogram of Magnetoencephalogram Background Noise Measurements.

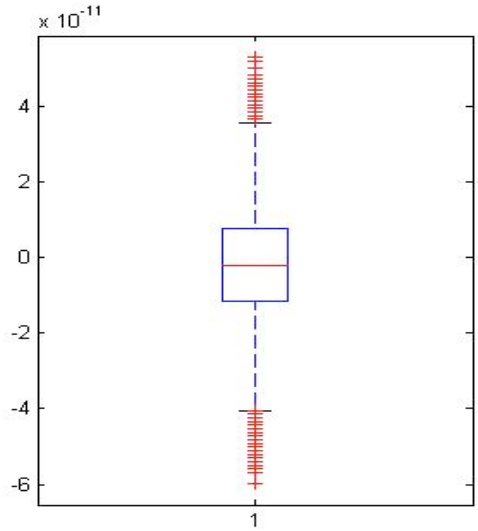


Figure 3F. Boxplot of Magnetoencephalogram Background Noise Measurements.

Tukey (1977) felt that to analyze any batch or sample of data, a statistician should know five-numbers. These are the minimum, 25th percentile, the median, the 75th percentile, and the maximum. These numbers constitute the **five-number summary**. The five-number summary for the MEG data is: the minimum is -5.9790e-11, the 25th percentile is -1.1572e-11, the median is -0.1929e-11, the 75th percentile is 0.7715e-11, and the maximum is 5.3040e-11.

E. The Gaussian Probability Model

Given the shape of the histogram and the information from the boxplot and the five-number summary, we will consider a Gaussian probability model. This continuous probability density is defined by the function:

$$f(x) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\}. \tag{3.17}$$

N.B. Statisticians say normal distribution for the Gaussian. Given this function, it suffices to specify this probability density to define the mean μ and the variance σ^2. The MEG noise distribution has $\mu =$ 0.0 and $\sigma = 1.4 \cdot 10^{-11}$. We interpret σ as the average amount of deviation away from the mean that we would expect to see. From Figure 3E we would not be surprised to see a noise as small as –1.4e-11 or as large as 1.4e-11 $(\mu - \sigma \le X \le \mu + \sigma)$, but we would consider it somewhat unusual to see noise smaller than –2.8e-11 or larger than 2.8e-11 $(X \le \mu - 2\sigma \text{ or } X \ge \mu + 2\sigma)$. We have $\Pr(X \le \mu - 2\sigma) = (X \ge \mu + 2\sigma) = 0.025$. That is, noise in this range would occur only 1 in 40 measurements.

We will not make use of the formula in Eq. 3.17 for the Gaussian probability model with mean μ and standard deviation σ because it is particularly difficult to evaluate since its integral cannot be obtained in closed form. There are two approximations that are worth remembering. These are

$$\Pr(\mu-\sigma \le X \le \mu+\sigma) = 0.67 \tag{3.18}$$

$$\Pr(\mu-2\sigma \le X \le \mu+2\sigma) = 0.95 \tag{3.19}$$

The actual value is 1.96 rather than 2. The number two is easier to remember. Kass (2006) refers to Eqs. 3.18 and 3.19 as the "$2/3$ to 0.95 Rule." In addition, $\Pr(\mu-3\sigma \le X \le \mu+3\sigma) \approx 0.997.$

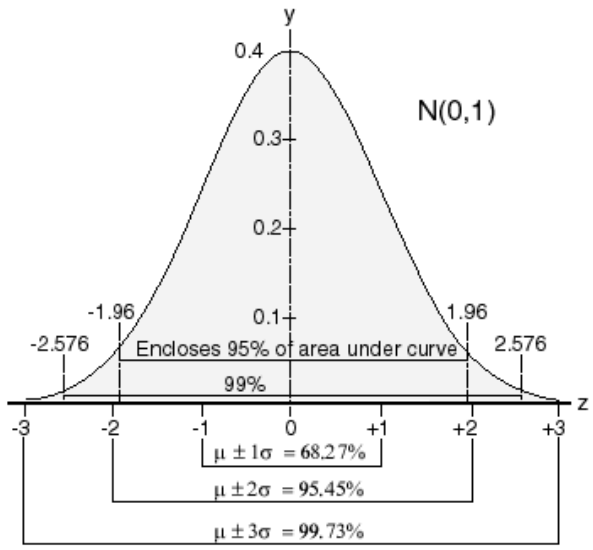


Figure 3G. Standard Gaussian probability density with mean $\mu = 0$ and variance $\sigma^2 = 1$.

The Gaussian distribution with mean $\mu = 0$ and $\sigma = 1$ is called the standard normal or Gaussian distribution. It is the one that is usually tabulated in textbooks because any Gaussian distribution can be transformed into it. To see this, we have from Eq. 3.15 if $X:\ N(\mu,\sigma^2)$ and $Y = aX + b$ then $Y:\ N(a\mu+b, a^2\sigma^2)$. As a special case, if $X:\ N(\mu,\sigma^2)$ and $Z = (X-\mu)/\sigma$ then $Z:\ N(0,1)$ is the standard Gaussian distribution. Therefore,

$$\begin{aligned}\Pr(a \le X \le b) &= \Pr\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) \\ &= \Pr\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right).\end{aligned} \tag{3.20}$$

The computation for any Gaussian probability density can be carried out in terms of the standard Gaussian distribution. This is how most statistical packages perform these computations. For the standard Gaussian the cdf is denoted as

$$\Phi(x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} \exp\{-\frac{1}{2}u^2\}du. \tag{3.21}$$

F. Properties of the Gaussian Distribution

For a Gaussian random variable we compute the expectation and variance as follows:

Expectation

$$E[X] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/2\sigma^2} dx .$$

Write $x = (x-\mu)+\mu$

$$E[X] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu) e^{-(x-\mu)^2/2\sigma^2} dx$$

$$+\mu \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx.$$

Let $y = x-\mu$

$$E[X] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} y e^{-y^2/2\sigma^2} dy + \mu \int_{-\infty}^{\infty} f(x) dx$$

$$= 0 + \mu \times 1 = \mu$$

Variance

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

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-(x-\mu)^2/2\sigma^2} dx.$$

Let $y = (x-\mu)/\sigma$ then

$$Var(X) = \frac{\sigma^2}{\sqrt{2\pi}} \int_{\infty}^{\infty} y^2 e^{-y^2/2} dy$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \sqrt{2\pi} = \sigma^2.$$

G. Quantile-Quantile Plots.

The Quantile-Quantile or Q-Q plot is a simple and quite general way to check whether a batch of data conforms, approximately to a particular probability model. The procedure compares the observed quantiles of the data to the theoretical quantiles of the probability distribution.

To compute the observed quantiles in a batch of observations we first put the data in ascending order according to the size of each observation: we rewrite as $y_{(1)}, y_{(2)}, \ldots, y_{(n)}$ where $y_{(1)}$ is the smallest value and $y_{(n)}$ is the largest. We then say that $y_{(r)}$ is the quantile of order $r/(n+1)$ (also called the $100r/(n+1)$ percentile). In **Section 3B,** we defined the quantiles of a probability distribution. If we are interested in the possibility that the data are a sample from a distribution having a cdf $F(x)$ then we may define the set of quantile

$$\eta_{(r)} = F^{-1}(r/(n+1)) \tag{3.22}$$

for $r = 1,\ldots,n$ and plot the ordered data against these values. That is, we plot the points $(y_{(1)},\eta_{(1)}),\ldots,(y_{(n)},\eta_{(n)})$. Such a plot is called a **Quantile-Quantile plot** or **Q-Q plot**. If the points fall roughly on the line $y = x$ then, we may be comfortable assuming that the data are a sample from a distribution having cdf $F(x)$. The decision of precisely which theoretical quantiles to pick in making this plot is actually somewhat arbitrary. An alternative is to take $\eta_{(r)} = F^{-1}(\frac{r-0.5}{n})$ and this is what we typically use.

Example 3.2 Magnetoencephalography (continued). We take $F(x) = \Phi(x)$ to be the cdf of the standard normal distribution, so we transform the data to $y_i = (x_i - \bar{x})/\hat{\sigma}$ where $\bar{x} = n^{-1}\sum_{i=1}^{n} x_i$ is the sample mean and $\hat{\sigma} = [n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2]^{1/2}$ is the sample standard deviation. A Q-Q plot of the data shown in Figure 3E is given in Figure 3H. Apart from small fluctuations partly due to rounding in the data, the plot follows closely the line $y = x$ suggesting that they agree very closely with a Gaussian distribution with mean $\hat{\mu} = \bar{x}$ and variance $\hat{\sigma}^2$.

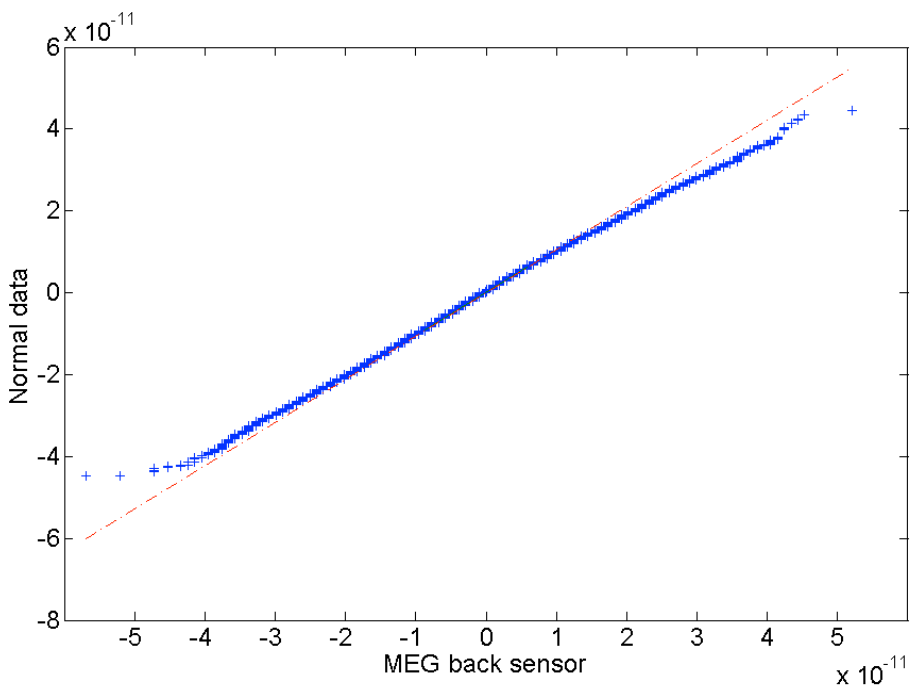


Figure 3H. Q-Q Plot of Magnetoencephalogram Background Noise Measurements.

H. Other Continuous Probability Models

There are other continuous probability models that we will use in the course. We list them and summarize their essential properties below.

1. Exponential Distribution

The probability density of the exponential distribution is

$$f(x) = \lambda e^{-\lambda x}. \qquad x > 0, \lambda > 0 \tag{3.23}$$

The exponential density is used to describe waiting time processes and inter-events times for systems that have no memory (Figure 3l, red curve). This model describes the inter-event occurrences for a Poisson process with rate λ. One example in neuroscience where the exponential probability model has been shown to hold is in the analysis of sleep bouts of mice. The sleep bouts, in contrast to the wake bouts, are well described by an exponential probability model (Lo et al. 2002). Another neuroscience example is given in Figure 2.10 on page 52 of Rice (2007). It shows several plots of the open time probability densities of nicotinic receptor channels as a function of different concentrations of the neuromuscular blocking agent suxamethonium (succinylcholine).

The expected value of an exponential random variable is

$$E[X] = \int_0^\infty x\lambda e^{-\lambda x} dx = \lambda^{-1} \tag{3.24}$$

and the variance is

$$\begin{aligned} Var[X] &= \int_0^\infty (x - \lambda^{-1})^2 \lambda e^{-\lambda x} dx \\ &= \int_0^\infty x^2 \lambda e^{-\lambda x} dx - \lambda^{-2} \\ &= 2\lambda^{-2} - \lambda^{-2} = \lambda^{-2}. \end{aligned} \tag{3.25}$$

2. Gamma Distribution

The probability density of the gamma distribution is

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad x > 0, \alpha > 0, \beta > 0 \tag{3.26}$$

where $\Gamma(\alpha)$ is the gamma function defined by $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$ The expected value is

$$\begin{aligned} E(X) &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x x^{\alpha-1} e^{-\beta x} dx \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha+1)}{\beta^{\alpha+1}} \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \times \frac{\alpha\Gamma(\alpha)}{\beta^{\alpha+1}} \\ &= \frac{\alpha}{\beta}. \end{aligned} \tag{3.27}$$

The variance is

$$Var(X) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{\infty} x^2 x^{\alpha-1} e^{-\beta\lambda x} dx - \left(\frac{\alpha}{\beta}\right)^2$$
$$= \frac{\alpha(\alpha+1)}{\beta^2} - \frac{\alpha^2}{\beta^2} = \frac{\alpha}{\beta^2}. \tag{3.28}$$

The gamma is a flexible family of probability densities for describing waiting time distributions (Figure 3I). The exponential distribution is a special case of the gamma where $\alpha = 1$ and $\beta = \lambda$. Since the exponential probability density is the gamma probability density with $\alpha = 1$, it follows that for the exponential random variable

$$Var(X) = \frac{1}{\beta^2}. \tag{3.29}$$

The chi-squared distribution with n degrees of freedom is the gamma probability density with $\alpha = n/2$ and $\beta = 1/2$.

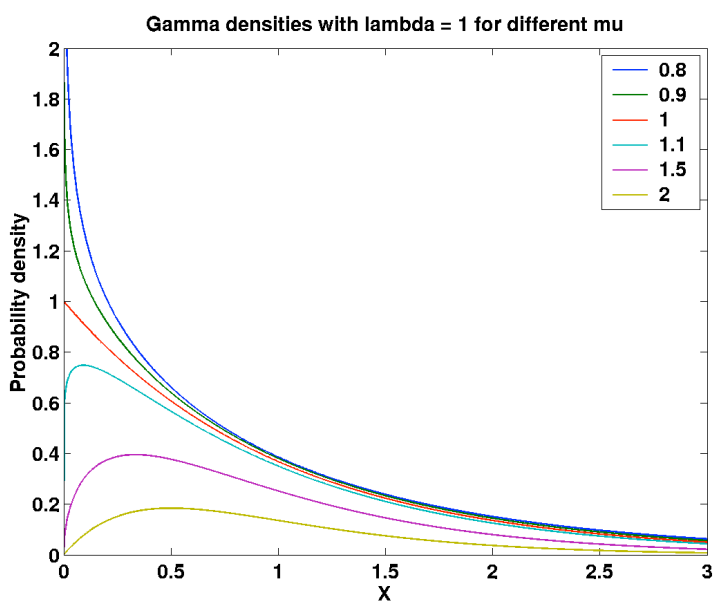


Figure 3I. Examples of Gamma probability densities.

3. Beta Distribution

The probability density of the beta distribution is

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \qquad 0 < x < 1,\ \alpha > 0,\ \beta > 0 \tag{3.30}$$

The expected value and the variance are respectively

$$E(X) = \frac{\alpha}{\alpha+\beta}$$
$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

To show that Eq. 3.30 defines a probability density and to verify the expressions for the expected value and the variance we wish to establish that

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

We can write

$$\Gamma(\alpha)\Gamma(\beta) = \int_0^\infty \int_0^\infty u^{\alpha-1}v^{\beta-1}e^{-(u+v)}dudv$$
$$= .\int_0^\infty u^{\alpha-1}e^{-u}du\int_0^\infty v^{\beta-1}e^{-u}dv.$$

We make the change of variables

$$x = u(u+v)^{-1}$$
$$y = u+v$$

We have

$$u = xy$$
$$v = (1-x)y$$

and that the Jacobian is y Therefore, we have

$$\begin{aligned}\Gamma(\alpha)\Gamma(\beta) &= \int_0^1\int_0^\infty x^{\alpha-1}(1-x)^{\beta-1}y^{\alpha+\beta-1}e^{-y}dydx \\ &= \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx\int_0^\infty y^{\alpha+\beta-1}e^{-y}dy \\ &= \Gamma(\alpha+\beta).\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx.\end{aligned} \tag{3.31}$$

We can use this result to establish the expectation and variance of the beta distribution. The beta distribution provides a highly flexible probability model for processes that are restricted to the interval $(0,1)$ (Figure 3J). In our Bayesian analysis it will serve as a prior probability density for the binomial p.

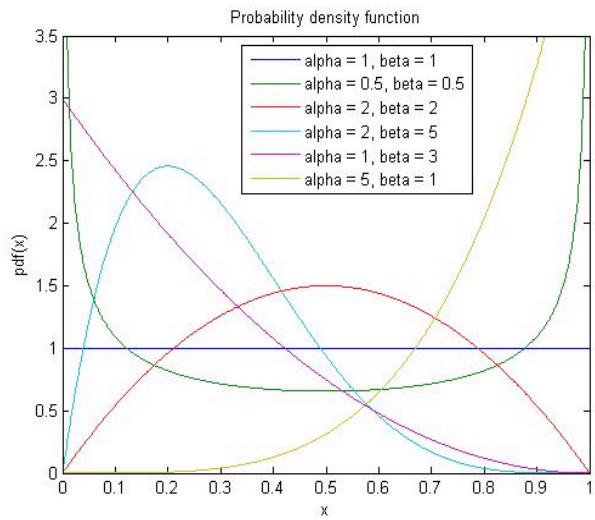


Figure 3J. Examples of Beta probability densities.

4. Inverse Gaussian Distribution

For the inverse Gaussian distribution the probability density, mean and variance are

$$
\begin{aligned}
&f(x) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\{-\tfrac{1}{2}\frac{\lambda(x-\mu)^2}{\mu^2 x}\} \\
&E(X) = \mu \\
&Var(X) = \mu^3 \lambda^{-1}.
\end{aligned}
\tag{3.32}
$$

Equation 3.32 defines the inverse Gaussian distribution for $\mu > 0$ and $\lambda > 0$, where μ is the mean and λ is the scale parameter. Schrödinger in (1915) gave one of the first derivations of this probability model and later it was used in neuroscience by Gerstein and Mandelbrot (1964). Gerstein and Mandelbrot proposed to analyze interspike interval distributions because the inverse Gaussian distribution is the probability model obtained from an elementary stochastic integrate-and-fire model (Tuckwell, 1988). The probability density is right-skewed (Figure 3K). If $(\mu/\lambda)^{-1}$ is large then, the inverse Gaussian probability density is well approximated by a Gaussian probability density. If we construct even the simplest stochastic dynamical models of neurons then, the resulting probability model for the interspike interval distribution is right-skewed and non-exponential. Hence, exponential waiting times or Poisson models are not the elementary models most likely to agree with neural spike train data. A time-dependent version of this model provides a great description of human heart beat dynamics (Barbieri et al. 2005; Barbieri and Brown, 2006).

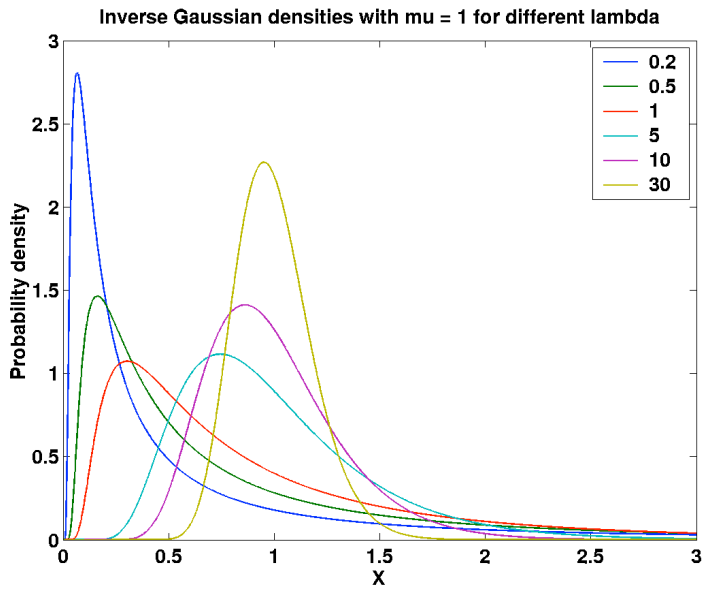


Figure 3K. Examples of inverse Gaussian probability densities.

III. Summary

We have developed the concept of continuous random variables and continuous probability models. We studied the Gaussian distribution, Finally, we reviewed some other useful continuous probability models for neuroscience. Together, **Lecture 2** and **Lecture 3** have established a set of probability models for data analysis.

Acknowledgments

I am grateful to Julie Scott for technical assistance, Uri Eden for preparing the figures and Jim Mutch and Sri Sarma for helpful editorial comments. Several sections of text for this lecture have been taken with permission from the class notes Statistical Methods in Neuroscience and Psychology by Robert Kass at Carnegie Mellon University, Statistics Department.

Textbook References

DeGroot MH, Schervish MJ. *Probability and Statistics*, 3rd edition. Boston, MA: Addison Wesley, 2002.

Rice JA, *Mathematical Statistics and Data Analysis*, 3rd edition. Belmont, CA: Duxbury, 2007.

Tukey JW. *Exploratory Data Analysis.* Reading, MA: Addison-Wesley, 1977.

Tuckwell, HC. *Introduction to Theoretical Neurobiology: Nonlinear and Stochastic Theories*, Volume 2. Cambridge: Cambridge University Press, 1988.

Literature References

Barbieri R, Matten EC, Alabi A, Brown EN. A point process model of human heart beat intervals: new definitions of heart rate and heart rate variability. *American Journal of Physiology: Heart and Circulatory Physiology* 2005, 288: H424-H435.

Barbieri R, Brown EN. Analysis of heart beat dynamics by point process adaptive filtering. *IEEE Transactions on Biomedical Engineering* 2006, 53(1): 4-12.

Lo CC, Nunes Amaral LA, Havlin S, Ch. Ivanov P, Penzel T, Peter JH, Stanley HE. Dynamics of sleep-wake transitions during sleep. *Europhys. Lett*. 2002, 57, 625-631.

Gerstein GL, Mandelbrot M. Random walk models for the spike activity of a single neuron. *Biophysical Journal* 4: 41-68, 1964.

Schrödinger E. Zur theori der fall und steigversuche an teilchen mit Brownscher bewegung. *Phys. Ze*., 16:289-295, 1915.

Website Reference

Kass RE (2006). Chapter 2: Probability and Random Variables
http://lib.stat.cmu.edu/~kass/smn/notes/2.pdf