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

Lecture 4: Transformations of Random Variables, Joint Distributions of Random Variables

I. Objectives

A. **Understand the basic rules for computing the distribution of a function of a random variable.**

B. **Understand how some important probability densities are derived using this method.**

C. **Understand the concept of the joint distribution of random variables.**

D. **Understand the bivariate Gaussian distribution.**

II. Transformations of Random Variables

If we assume X is a continuous random variable with pdf $f(x)$ we consider the following transformations of X. They are

1. $Y = aX + b$ (Linear)
2. $Y = |X|$ (Absolute Value) (4.1)
3. $Y = X^2$ (Quadratic)
4. $Y = g(x)$ (Monotonic)

In each case, we will want to find $f_y(y)$.

Example 4.1 Linear Transformation. Take $Y = aX + b$. Find $f_y(y)$ (Rice: pp 64-67, Problem 62). Find the cdf and the pdf of Y.

$$\begin{aligned} F_y(y) &= \Pr(Y \le y) \\ &= \Pr(aX + b \le y) \\ &= \Pr(X \le \frac{y-b}{a}) \\ &= F_x(\frac{y-b}{a}) = \int_{-\infty}^{\frac{y-b}{a}} f_x(x)dx \end{aligned} \tag{4.2}$$

which is the cdf of Y. We find the pdf $f_y(y)$ by differentiating $F_y(y)$ and considering two cases

Case i $a > 0$:

$$\begin{aligned} f_y(y) &= \frac{d}{dy}(F_x(\frac{y-b}{a})) \\ &= f_x(\frac{y-b}{a}) \times \frac{1}{a} \end{aligned} \tag{4.3}$$

Case ii $a < 0$:

$$\begin{aligned} f_y(y) &= \frac{d}{dy}(F_x(\frac{y-b}{a})) \\ &= f_x(\frac{y-b}{a}) \times \frac{1}{a} < 0 \end{aligned}$$

for all y and hence, it is not a probability density. Thus, we multiply by $\frac{1}{-a}$ if $a < 0$. Putting these two cases together we have

$$f_y(y) = f_x(\frac{y-b}{a})\frac{1}{|a|} \tag{4.4}$$

Remark 4.1. Example 4.1 illustrates a special case of the general result we establish in **Proposition 4.1**.

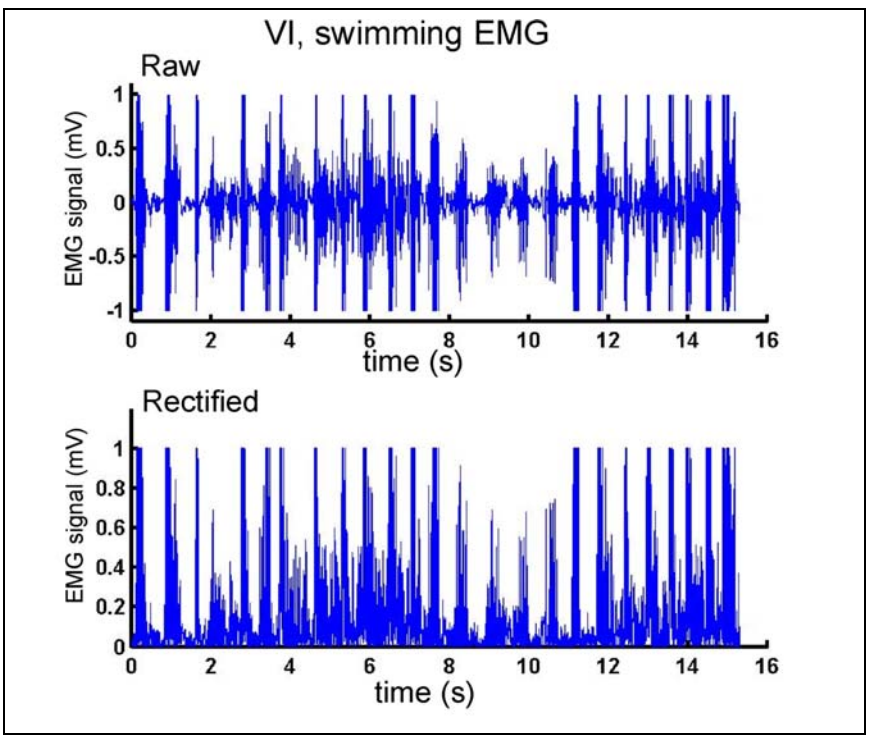

Figure 4A. Fifteen-second time course of raw electromyogram (EMG) signals (panel 1) and the rectified EMG signals recorded from the semitendonosis muscle of a frog executing swimming motions.

Example 4.2. Probability Density of a Rectified Electromyogram Signal. $X \sim N(\mu, \sigma^2)$ $Y = |X|$ Find $f_y(y)$. This is the problem that must be considered in analyzing electromyographic (EMG) data. An EMG is a recording of the electrical impulses transmitted through a group of muscle fibers. The impulses and the electrical potentials are both positive and negative (Figure 4A, panel 1). Because the force that is generated by a muscle is only positive, the data are analyzed after rectification. That is taking the absolute value of each observation (Figure 4B, panel 2). If we assume that the data are Gaussian, then we need to compute the probability density of the absolute value of a Gaussian random variable.

$$\begin{aligned}
\Pr(Y \le y) &= \Pr(|X| \le y) \\
&= \Pr(-y \le X \le y) \\
&= \Pr\left(\frac{-y-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{y-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{y-\mu}{\sigma}\right) - \Phi\left(\frac{-y-\mu}{\sigma}\right) \\
&= \left(\frac{1}{2\pi}\right)^{\frac{1}{2}}\left[\int_{-\infty}^{\frac{y-\mu}{\sigma}} \exp\{-\frac{1}{2}x^2\}dx - \int_{-\infty}^{\frac{-y-\mu}{\sigma}} \exp\{-\frac{1}{2}x^2\}dx\right]
\end{aligned} \tag{4.5}$$

Hence,

$$\begin{aligned}
f_Y(y) &= \phi\left(\frac{y-\mu}{\sigma}\right)\frac{1}{\sigma} + \phi\left(\frac{-y-\mu}{\sigma}\right)\frac{1}{\sigma} \\
&= \frac{1}{\sigma}\left[\phi\left(\frac{y-\mu}{\sigma}\right) + \phi\left(\frac{-y-\mu}{\sigma}\right)\right].
\end{aligned} \tag{4.6}$$

If $\mu = 0$ then, the probability density is $f_y(y) = \frac{2}{\sigma}\phi\left(\frac{y}{\sigma}\right)$ $y > 0$. This is <u>not</u> the probability density of the sum of two Gaussian random variables. This we will study in **Lectures 5** and **6**.

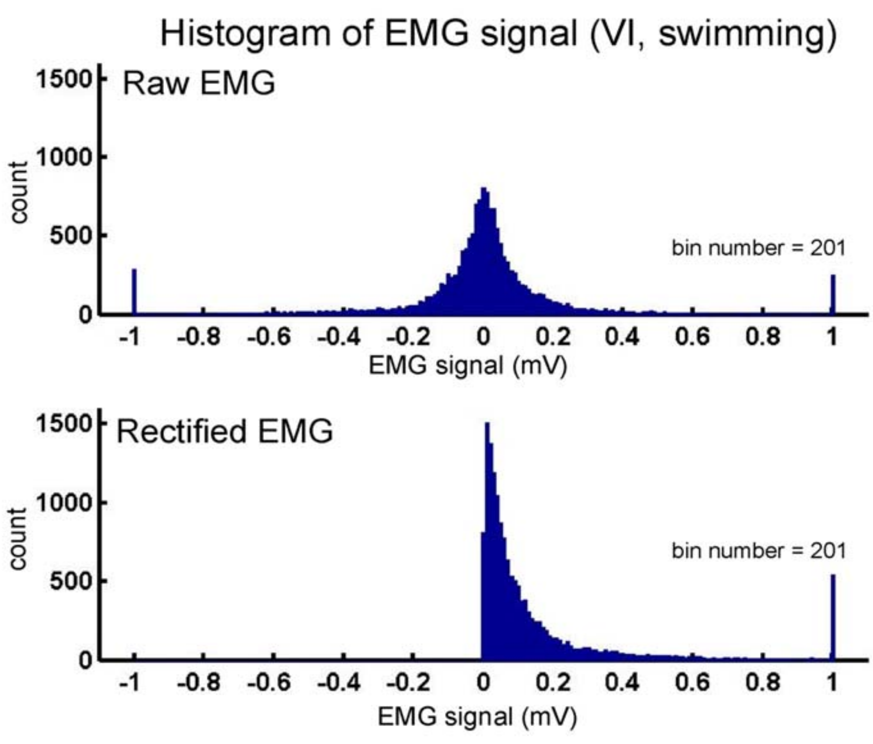

Figure 4B. Histograms of the raw EMG signals (panel 1) and the rectified EMG signals (panel 2) from Figure 4A.

Example 4.2 (continued). Figure 4B shows the histogram of the raw EMG signal (panel 1) and the rectified EMG signal (panel 2). How can we evaluate if the Gaussian assumption is reasonable?

Example 4.3. If $X \sim N(0,1)$ and $Y = X^2$ Find $f_y(y)$.

We have

$$\begin{aligned}\Pr(Y \le y) &= \Pr(X^2 \le y) \\ &= \Pr(-y^{\frac{1}{2}} \le X \le y^{\frac{1}{2}}) \\ &= \Phi(y^{\frac{1}{2}}) - \Phi(-y^{\frac{1}{2}})\end{aligned} \tag{4.7}$$

On differentiating we obtain

$$\begin{aligned}f_y(y) &= \frac{1}{2} y^{-\frac{1}{2}}[\phi(y^{\frac{1}{2}}) + \phi(-y^{\frac{1}{2}})] \\ &= y^{-\frac{1}{2}}\phi(y^{\frac{1}{2}}) \\ &= (\frac{1}{2\pi})^{\frac{1}{2}} y^{-\frac{1}{2}} \exp\{-\frac{1}{2}y\} \\ &= \frac{y^{-\frac{1}{2}}}{\Gamma(\frac{1}{2})}\exp\{-\frac{1}{2}y\}\end{aligned} \tag{4.8}$$

which is a gamma density with $\alpha = \frac{1}{2}$ and $\beta = \frac{1}{2}$. This density is called a chi-squared density with 1 degree of freedom written χ_1^2. The χ^2 distribution with n degrees of freedom will be important in our data analyses.

We can state a generalization of what we have done in computing the distribution of transformations of random variables.

Proposition 4.1. If X is a continuous random variable with pdf $f_x(x)$ and let $Y = g(X)$ where g is a differentiable, strictly monotonic function on some interval I. Suppose $f(x) = 0$ if $x \notin I$. Then Y has the density function

$$f_y(y) = f_x(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right| \tag{4.9}$$

Proof: g is strictly monotonic means that if $x_1 < x_2$ then $g(x_1) < g(x_2)$ if g is increasing and $g(x_2) < g(x_1)$ if g is decreasing. On the interval I the mapping from X to Y is one-to-one. Hence, g^{-1} exists, and

$$\begin{aligned}\Pr(Y \le y) &= \Pr(g(X) \le y) \\ &= \Pr(X \le g^{-1}(y)) \\ &= F_x(g^{-1}(y)) \\ f_y(y) &= F_x^{'}(g^{-1}(y)) \cdot \frac{d}{dy}(g^{-1}(y)) \\ &= f_x(g^{-1}(y)) \cdot \frac{d}{dy}(g^{-1}(y))\end{aligned} \tag{4.10}$$

If $\frac{d}{dy}(g^{-1}(y)) < 0,$ then we multiply by $-\frac{d}{dy}(g^{-1}(y))$ to insure that $f_y(y) \ge 0$ for all $y.$ Because g is a one-to-one mapping, we can write a more intuitive derivation because we must have, working with differentials,

$$\begin{aligned}f_y(y)dy &= f_x(x)dx \\ f_y(y) &= f_x(x)\frac{dx}{dy}\end{aligned} \tag{4.11}$$

Now $x = g^{-1}(y)$ so that

$$f_y(y) = f_x(g^{-1}(y)) \cdot \frac{d}{dy}(g^{-1}(y)) \tag{4.12}$$

Again, we note that we take the absolute value of the derivative to make sure that $f_y(y) \ge 0$ for all $y.$

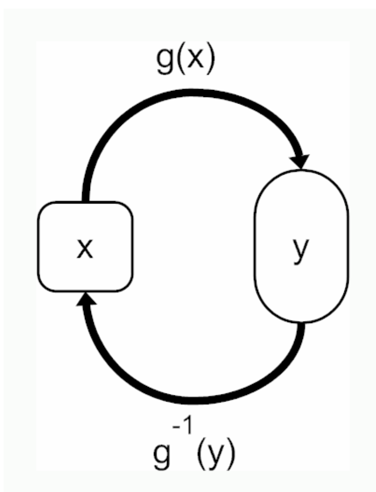

Figure 4C. Mapping and inverse mapping between x and y.

II. Joint Distributions of Random Variables

We consider now joint distributions of random variables. Being able to analyze the joint distribution of n random variables that represent a collection of data will be essential for our statistical modeling. In general, the behavior of two random variables X and Y is given by the **joint cumulative distribution function**

$$F(x,y) = \Pr(X \le x,\ Y \le y) \tag{4.13}$$

whether X and Y are continuous or discrete.

A. Discrete Random Variables

If X and Y are discrete random variables defined on the same outcome space and they assume values respectively $x_1, x_2, \ldots,$ and $y_1, y_2, \ldots,$ then their **joint probability mass function** is

$$p(x_i, y_j) = \Pr(X = x_i, Y = y_j). \tag{4.14}$$

We have that $\sum_i \sum_j p(x_i y_j) = 1$ so that the joint pmf is well defined. We illustrate this idea with a simple example.

Example 2.0 (continued). One roll of two fair dice yields

$$\begin{array}{ccccccc} & & & Y & & & \\ & 1,1 & 1,2 & 1,3 & 1,4 & 1,5 & 1,6 \\ & 2,1 & 2,2 & 2,3 & 2,4 & 2,5 & 2,6 \\ X & 3,1 & 3,2 & 3,3 & 3,4 & 3,5 & 3,6 \\ & 4,1 & 4,2 & 4,3 & 4,4 & 4,5 & 4,6 \\ & 5,1 & 5,2 & 5,3 & 5,4 & 5,5 & 5,6 \\ & 6,1 & 6,2 & 6,3 & 6,4 & 6,5 & 6,6 \end{array}$$

where X is the value on the face of the first die and Y is the value on the face of the second die. We have

$$p(x_i, y_i) = \frac{1}{36} \tag{4.15}$$

for $x_i = 1,2,3,4,5,6$ and $y_i = 1,2,3,4,5,6$ and 0 otherwise.

For example

$$p(3,1) = p(2,2) = p(1,3) = \frac{1}{36} \tag{4.15}$$

and

$$\Pr(X = 1) = \sum_{i=1}^{6} p(X = 1, Y_i) = p(X = 1, Y = 1) + p(X = 1, Y = 2) + p(X = 1, Y = 3) + p(X = 1, Y = 4) + p(X = 1, Y = 5)$$
$$+ p(X = 1, Y = 6) = \frac{1}{6}.$$

(4.17)

To find the pmf of X we sum across the columns, whereas to find the pmf of $Y,$ we sum across the rows. This defines the **marginal probability mass functions** of X and Y respectively as

$$p_x(x) = \sum_i p(x, y_i)$$
$$p_y(y) = \sum_i p(x_i, y) \tag{4.18}$$

B. Continuous Random Variables

If X and Y are continuous random variables with a joint cdf $F(x, y),$ their **joint probability density function** is the piecewise continuous function of two variables $f(x, y)$ defined as

$$f(x, y) \ge 0$$
$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1. \tag{4.19}$$

For any “reasonable” set A

$$\Pr(X, Y) \in A) = \iint_A f(x, y)dxdy \tag{4.20}$$

If $A = \{(x, y) \mid X \le x, Y \le y\},$

$$F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u, v)dudv. \tag{4.21}$$

From the fundamental theorem of multivariate calculus, it follows that

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} \tag{4.22}$$

wherever the derivative is defined.

For small δx and $\delta y,$ if f is continuous, we have

$$\Pr(x \le X \le x + \delta x,\ y \le Y \le y + \delta y) = \int_{x}^{x+\delta x}\int_{y}^{y+\delta y} f(u, v)dvdu \approx f(x, y)dxdy. \tag{4.23}$$

Roughly speaking for a 1-dimensional continuous random variable computing probabilities corresponds to computing the area under a curve, whereas for 2-dimensional continuous random variables, computing probabilities corresponds to computing the volume under a surface.

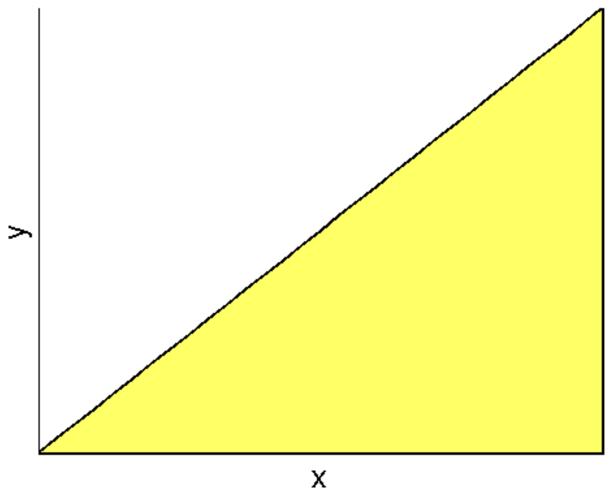

Figure 4.D. The domain of integration for Example 4.4 is the region $\{(x,y) \mid 0 \le y \le x \le 1\}$.

Example 4.4 (Rice, pp. 75-76). Consider the bivariate density

$$f(x,y) = \frac{12}{7}(x^2 + xy) \qquad 0 \le x \le 1 \qquad 0 \le y \le 1 \tag{4.24}$$

Compute

$$\Pr(X > Y) \tag{4.25}$$

This is the set

$$\{(x,y) \mid 0 \le y \le x \le 1\} \tag{4.26}$$

$$\begin{aligned} \Pr(X > Y) &= \frac{12}{7}\int_0^1\int_0^x (x^2 + xy)dydx \\ &= \frac{12}{7}\int_0^1 (x^2 y + \frac{xy^2}{2}\Big|_0^x)dx = \frac{12}{7}\int_0^1 (x^3 + \frac{x^3}{2})dx \\ &= \frac{36}{14}\int_0^1 x^3 dx = \frac{36}{14}\frac{x^4}{4}\Big|_0^1 = \frac{9}{14} \end{aligned} \tag{4.27}$$

The **marginal cumulative distribution function** of X is

$$\begin{aligned} F_X(x) &= \Pr(X \le x) \\ &= \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(u,y)dydu \end{aligned} \tag{4.28}$$

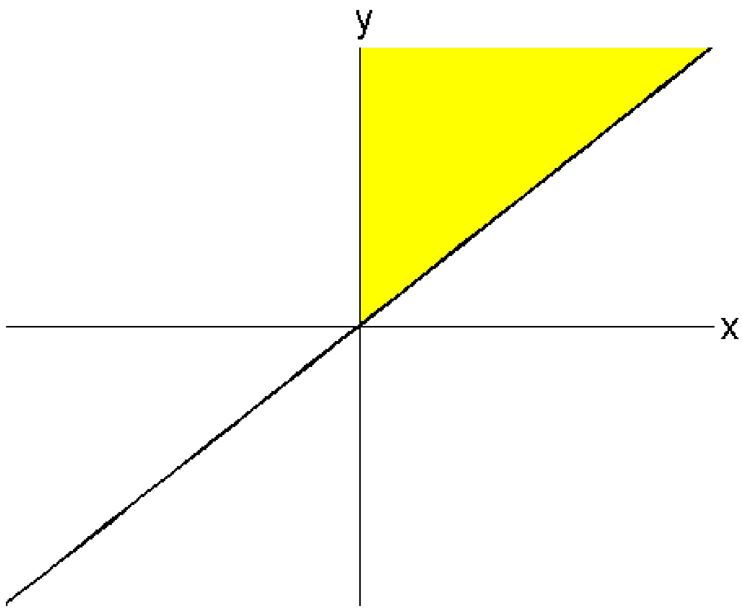

Figure 4.E. The domain of integration for Example 4.5 is the region $0 \le x \le y$ **.**

Example 4.5 (Rice, pp. 79-80). Consider the joint probability density

$$f(x,y) = \begin{cases} \lambda^2 e^{-\lambda y} & 0 \le x \le y,\ \lambda > 0 \\ 0 & \end{cases} \tag{4.29}$$

Find the marginal pdf of X and the marginal pdf of Y.

For $f_x(x)$ we have

$$\begin{aligned} f_x(x) &= \int_{-\infty}^{\infty} f_{xy}(x,y)dy \\ &= \int_x^{\infty} \lambda^2 e^{-\lambda y} dy = \frac{\lim}{b \to \infty} - \lambda e^{-\lambda b} + \lambda e^{-\lambda x} = \lambda e^{-\lambda x} \end{aligned} \tag{4.30}$$

Hence, X has an exponential pdf. Now $f_{xy}(x,y) = 0$ for $x \le 0$ and $x > y$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{xy}(x)dx = \int_0^y \lambda^2 e^{-\lambda y} dx = \lambda^2 e^{-\lambda y}\, x|_0^y = \lambda^2 y e^{-\lambda y}. \tag{4.31}$$

Hence, Y has a gamma pdf with $\alpha = 2$ and $\beta = \lambda$.

C. Bivariate Gaussian Probability Density
An important bivariate probability density is given by

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y(1-\rho^2)^{\frac{1}{2}}} \times \exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}]\} \quad (4.32)$$

where $x \in (-\infty,\infty)$, $y \in (-\infty,\infty)$, $\mu_x \in (-\infty,\infty)$, $\mu_y \in (-\infty,\infty)$, $\sigma_x > 0, \sigma_y > 0$ and $-1 < \rho < 1$. We have that μ_x is the mean of X, μ_y is the mean of Y, σ_x is the standard deviation of X, σ_y is the standard deviation of y and ρ is the correlation coefficient. We have that $f(x,y)$ is constant if

$$\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} = c \quad (4.33)$$

The locus of these points is an ellipse centered at (μ_x, μ_y). If $\rho = 0$ the axes are parallel to the x and y axes. If $\rho \neq 0$, then axes are tilted. If $\rho > 0$, then the tilt is positive. If $\rho < 0$, then the tilt is negative.

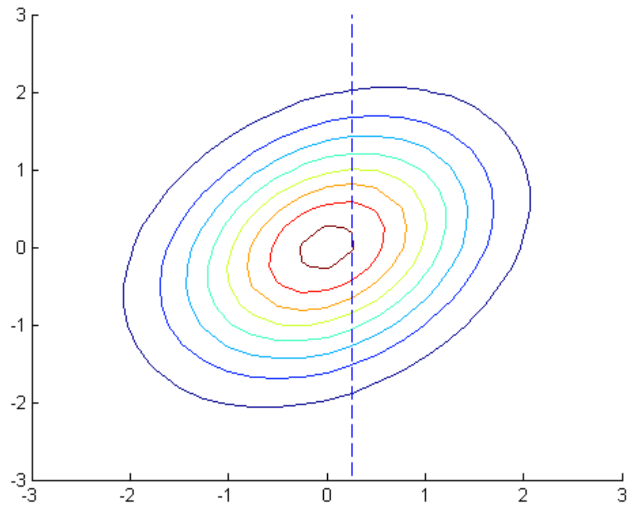

Figure 4.F. Bivariate Gaussian probability density.

The marginal distributions of x and y are respectively $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$. To show this we write

$$f_x(x) = \int_{-\infty}^{\infty} f_{xy}(x,y)dy \quad (4.34)$$

Making the change of variables $u = \frac{(x-\mu_x)}{\sigma_x}$ and $v = \frac{(y-\mu_y)}{\sigma_y}$ gives

$$f_x(x) = \frac{1}{2\pi\sigma_x(1-\rho^2)^{\frac{1}{2}}} \int_{-\infty}^{\infty} \exp\{-\frac{1}{2(1-\rho^2)}(u^2 + v^2 - 2\rho uv)\}dv \quad (4.35)$$

Complete the square in v

$$\begin{aligned} u^2 + v^2 - 2\rho uv &= u^2 + v^2 - 2\rho uv + \rho^2 u^2 - \rho^2 u^2 \\ &= v^2 - 2\rho uv + \rho^2 u^2 + u^2 + \rho^2 u^2 = (v - \rho u)^2 + u^2(1 - \rho^2) \end{aligned} \tag{4.36}$$

Hence, substituting and using the definition of a Gaussian integral

$$\begin{aligned} f_x(x) &= \frac{1}{2\pi\sigma_x(1-\rho^2)^{\frac{1}{2}}} \exp\{-\frac{1}{2}u^2\} \int_{-\infty}^{\infty} \exp\{-\frac{1}{2}\frac{(v-\rho u)^2}{(1-\rho^2)} dv \\ &= \frac{1(2\pi(1-\rho^2))^{\frac{1}{2}}}{2\pi\sigma_x(1-\rho^2)^{\frac{1}{2}}} \exp\{-\frac{1}{2}u^2\} = (\frac{1}{2\pi\sigma_x^2})^{\frac{1}{2}} \exp\{-\frac{1}{2}u^2\} \\ &= (2\pi\sigma_x^2)^{-\frac{1}{2}} \exp\{-\frac{1}{2}\frac{(x-\mu_x)^2}{\sigma_x^2}\} \end{aligned} \tag{4.37}$$

Since $u = (x - \mu_x)/\sigma_x$.

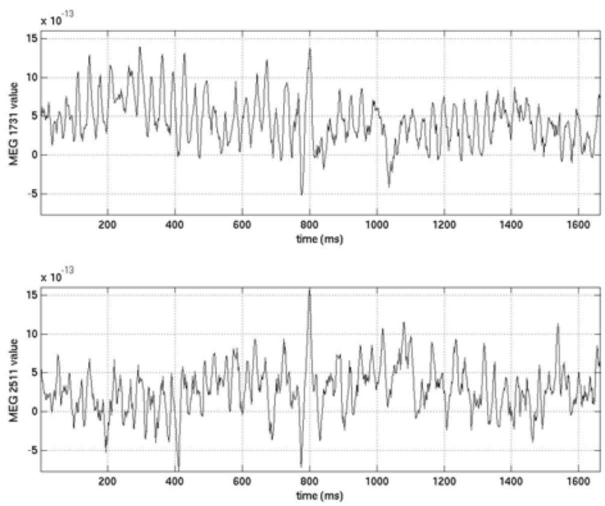

Figure 4G. 1.6 seconds of two time-series of simultaneous recordings from a front (panel 1) and a back (panel 2) magnetoencephalogram SQUID sensor.

Example 4.6. The Joint PDF of Two MEG Sensors. Figure 4H shows the scatterplot of simultaneous recordings from the two MEG SQUID sensors in Figure 4G. Do the data appear to be bivariate Gaussian? Histograms with superimposed estimates of the marginal probability densities are shown in panels 1 and 3 of Figure 4I The corresponding Q-Q (probability) plots for the two marginal probability densities are shown in panels 2 and 4 of Figure 4I. Do the marginal pdf's appear to be Gaussian? Is the correlation coefficient positive or negative for this probability density?

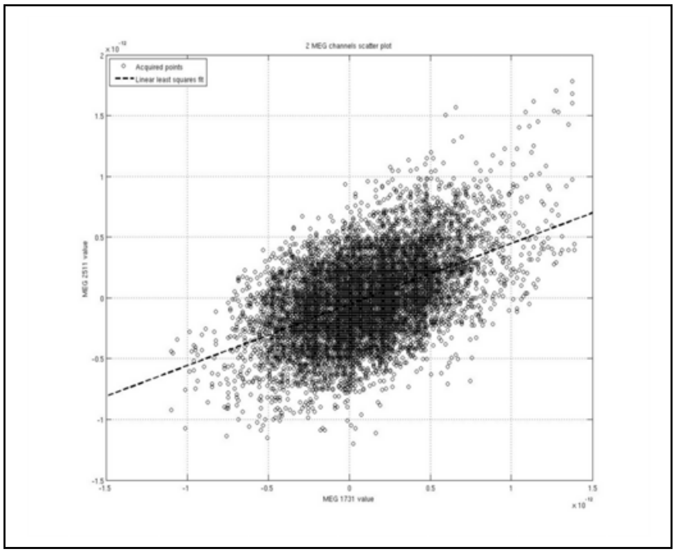

Figure 4H. Scatterplot of the simultaneous recordings from a front (x-axis) and a back (y-axis) magnetoencephalogram SQUID sensor.

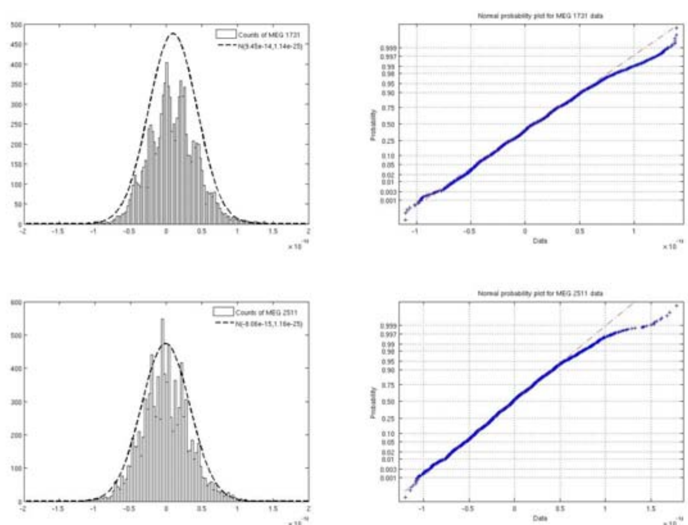

Figure 4I. Histogram (panel 1) and Gaussian probability density estimate (panel 1, dashed curve) of the front SQUID sensor measurements. Histogram (panel 3) and Gaussian probability density estimate (panel 3, dashed curve) of the back SQUID sensor measurements. Q-Q (Probability) plot of the front SQUID sensor measurements (panel 2). Q-Q (Probability) plot of the back SQUID sensor measurements (panel 4, blue dots). The dashed line in panels 2 and 4 show the line of exact agreement between the SQUID measurements and the estimated Gaussian density.

D. Independent Random Variables

Definition 4.1. The random variables $X_1, X_2, ..., X_n$ are **independent** if their joint cdf can be expressed as a product of their marginals. That is

$$F(X_1, X_2, ... X_n) = \prod_{i=1}^{n} F_{x_i}(X_i) \tag{4.38}$$

or equivalently, their joint pmf's for discrete random variables or their joint pdf's for continuous random variables may be

$$f(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} f_{x_i}(X_i). \tag{4.39}$$

Remark 4.2. For a bivariate Gaussian probability density, if $\rho = 0$ the joint probability density is

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi}\frac{1}{(\sigma_x^2\sigma_y^2)^{\frac{1}{2}}}\exp\{-\frac{1}{2}[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}]\} \\ &= (\frac{1}{2\pi\sigma_x^2})^{\frac{1}{2}}\exp\{-\frac{1}{2}[\frac{(x-\mu_x)^2}{\sigma_x^2}]\}(\frac{1}{2\pi\sigma_y^2})^{\frac{1}{2}}\exp\{-\frac{1}{2}(\frac{(y-\mu_y)^2}{\sigma_y^2})\} \\ &= f_x(x)f_y(y). \end{aligned} \tag{4.40}$$

and hence, we see the very important fact that if two Gaussian random variables are uncorrelated, or more generally, n Gaussian random variables are uncorrelated, then they are also independent. In general, if two random variables are uncorrelated, they are not independent. However, if two random variables are independent, they are uncorrelated. As stated in **Lecture 1**, independence is a strong condition. We will use an independence assumption frequently in constructing a joint probability density for the data in our statistical models.

IV. Summary

We have shown how to construct the probability densities for a transformation of a random variable. In addition, we have shown how to define joint probability densities for two or more random variables.

Acknowledgments. I am grateful to Julie Scott for technical assistance in preparing this lecture, to Emilio Bizzi and Vincent Cheung for the EMG data in Figures 4A and 4B to Matti Hämäläinen and Antonio Molins for the MEG data in Figures 4G-4I. I am grateful to Vincent Cheung for preparing Figures 4A and 4B , to Antonio Molins for preparing Figures 4G-4I, and to Paymon Hosseini for preparing Figures 4C- 4F.

Reference

Rice JA. *Mathematical Statistics and Data Analysis*, 3rd edition. Boston, MA, 2007.