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

Lecture 5: Conditional Distributions and Functions of Jointly Distributed Random Variables

I. Objectives

> **Understand the concept of a conditional distribution in the discrete and continuous cases.**
>
> **Understand how to derive the distribution of the sum of two random variables.**
>
> **Understand how to compute the distribution for the transformation of two or more random variables.**

II. Conditional Distributions

Just as we used conditional probabilities in **Lecture 1** to evaluate the likelihood of one event given another, we develop here the concepts of discrete and continuous conditional distributions and discrete and continuous conditional probability mass functions and probability density functions to evaluate the behavior of one random variable given knowledge of another.

A. Discrete Conditional Distributions

If X and Y are jointly distributed discrete random variables, the conditional probability that $X = x_i$ given $Y = y_j$ is

$$\begin{aligned}\Pr(X = x_i \mid Y = y_j) &= \frac{\Pr(X = x_i, Y = y_j)}{\Pr(Y = y_j)} \\ &= \frac{p_{xy}(x_i, y_j)}{p_y(y_j)}\end{aligned} \tag{5.1}$$

provided that $p_y(y_j) > 0$. This is the **conditional probability mass function** of X given $Y = y_j$. The numerator is a function of x for a fixed value of y. Similarly, we have

$$\Pr(Y = y_j \mid X = x_i) = \frac{p_{xy}(x_i, y_j)}{p_x(x_i)} \tag{5.2}$$

We also have

$$p_{xy}(x_i, y_j) = p_x(x_i)p_{y|x}(y_i \mid x_i) = p_y(y_j)p_{x|y}(x_i \mid y_j). \tag{5.3}$$

Example 5.1 Melencolia I. The following discrete joint probability mass function is based on a magical square in Albert Dürer's engraving Melencolia I.

X \ Y	1	2	3	4
1	$16/136$	$3/136$	$2/136$	$13/136$
2	$5/136$	$10/136$	$11/136$	$8/136$
3	$9/136$	$6/136$	$7/136$	$12/136$
4	$4/136$	$15/136$	$14/136$	$1/136$

(5.4)

Find $\Pr(Y=1 \mid X=3)$ and $\Pr(X=4 \mid Y=2)$

$$
\begin{aligned}
\Pr(Y=1 \mid X=3) &= \frac{\Pr(Y=1 \cap X=3)}{\Pr(X=3)} \\
&= \frac{\Pr(Y=1 \cap X=3)}{\sum_{i=1}^{4} \Pr(Y=y_i \cap X=3)} \\
&= \frac{9}{136} \Big/ \frac{1}{136}(9+6+7+12) = 9/34 \\
\Pr(X=4 \mid Y=2) &= \frac{\Pr(X=4 \cap Y=2)}{\sum_{i=1}^{4} \Pr(X=x_i \cap Y=2)} \\
&= \frac{15}{136} \Big/ \frac{1}{136}(3+10+6+15) \\
&= \frac{15}{34}
\end{aligned}
\tag{5.5}
$$

Example 5.2 Suppose that each acetylcholine molecule released at the neuromuscular junction of a frog has a probability p of being defective. If the distribution of the number of molecules released in a 10 msec time interval is a Poisson distribution with parameter λ, what is the distribution of the number of defective acetylcholine molecules?

Let N be the number of acetylcholine molecules released in 10 msec and let X be the number that is defective. If we assume that one molecule being defective is independent of any other being defective, we have that the conditional pmf of X given $N=n$ is the binomial distribution

$$
\Pr(X=k \mid N=n) = \binom{n}{k} p^k (1-p)^{n-k} \tag{5.6}
$$

And we have that

$$\Pr(N=n)=\frac{\lambda^n e^{-\lambda}}{n!}. \tag{5.7}$$

By the Law of Total Probability (**Lecture 1**), we have

$$\begin{aligned}
\Pr(X=k) &= \sum_{n=0}^{\infty}\Pr(N=n)\Pr(x=k\mid N=n)\\
&= \sum_{n=k}^{\infty}\frac{\lambda^n e^{-\lambda}}{n!}\binom{n}{k}p^k(1-p)^{n-k}\\
&= \sum_{n=k}^{\infty}\frac{\lambda^n e^{-\lambda}}{n!}\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}\\
&= \frac{(\lambda p)^k}{k!}e^{-\lambda}\sum_{n=k}^{\infty}\frac{\lambda^{n-k}(1-p)^{n-k}}{(n-k)!}\\
&= \frac{(\lambda p)^k}{k!}e^{-\lambda}e^{\lambda(1-p)}\\
&= \frac{(\lambda p)^k}{k!}e^{-\lambda p}
\end{aligned} \tag{5.8}$$

which is a Poisson model with rate parameter λp.

B. Continuous Conditional Distribution

If X and Y are continuous random variables with joint probability density $f_{xy}(x,y)$, then the **conditional probability density** of Y given X is defined as

$$f_{y|x}(y\mid x)=\frac{f_{xy}(x,y)}{f_x(x)} \tag{5.9}$$

for $0<f_x(x)<\infty$ and 0 otherwise. If we define this in terms of differentials we have

$$\begin{aligned}
\Pr(y<Y\le y+dy\mid x\le X\le x+dx) &= \frac{\Pr(y<Y\le y+dy \cap x\le X\le x+dx)}{\Pr(x\le X\le x+dx)}\\
&= \frac{f_{xy}(x,y)dxdy}{f_x(x)dx}\\
&= \frac{f_{xy}(x,y)dy}{f_x(x)}.
\end{aligned} \tag{5.10}$$

Hence, we have

$$f_{y|x}(y \mid x) = \lim_{dy \to \infty} \frac{(\Pr(y < Y \leq y + dy \mid x \leq X \leq x + dx)}{dy}$$
$$= \frac{f_{xy}(x, y)}{f_x(x)}$$

The numerator is a function of y for a fixed value of $x.$ Similarly, we can define

$$\begin{aligned} f_{x|y}(x \mid y) &= \frac{f_{xy}(x, y)}{f_y(y)} \\ f_{xy}(x, y) &= f_{y|x}(y \mid x) f_x(x) \end{aligned} \tag{5.11}$$

and the marginal probability density of Y is

$$f_Y(y) = \int f_{y|x}(y \mid x) f_x(x) dx \tag{5.12}$$

which is the **Law of Total Probability** for continuous random variables.

Example 4.6 (continued). Figure 4H shows the joint probability density of the two MEG sensors. Consider the distribution of the front sensor (y-axis) given that the value of the back sensor (x-axis) is 0.25. If the bivariate Gaussian probability density is a reasonable model for these data, then we can show that the conditional distribution of Y given X is the Gaussian probability density defined as

$$f_{Y|X}(y \mid x) = [2\pi\sigma_y^2(1-\rho^2)]^{-\frac{1}{2}} \times \exp\left\{ -\frac{1}{2} \frac{[y - \mu_y - \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x)]^2}{\sigma_y^2(1-\rho^2)} \right\} \tag{5.13}$$

We have

$$\begin{aligned} E(Y \mid X) &= \mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x) \\ Var(Y \mid X) &= \sigma_y^2(1-\rho^2). \end{aligned} \tag{5.14}$$

Note that if $\rho = 0,$ we have the marginal density of Y is the Gaussian density with $E(Y) = \mu_y$ and $Var(Y) = \sigma_y^2,$ consistent with the idea that X and Y would be independent. Notice also that $\sigma_y^2(1-\rho^2) \leq \sigma_y^2$ showing that knowledge about X reduces the uncertainty in $Y.$

The conditional mean $\mu_{y|x} = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x)$ is called the regression line of y on x. It gives the "best" prediction of y given x. When we study regression analysis we will make this statement precise.

Example 4.5 (continued). In this example we have the joint and marginal densities as

$$
\begin{aligned}
f_{xy}(x,y) &= \lambda^2 e^{-\lambda y} && 0 \le x \le y \\
f_x(x) &= \lambda e^{-\lambda x} && 0 \le x \\
f_y(y) &= \lambda^2 y e^{-\lambda y} && 0 \le y
\end{aligned}
\tag{5.15}
$$

The conditional densities are

$$f_{y|x}(y \mid x) = \frac{f_{xy}(x,y)}{f_x(x)} = \frac{\lambda^2 e^{-\lambda y}}{\lambda e^{-\lambda x}} = \lambda e^{-\lambda(y-x)} \tag{5.16}$$

for $0 < x < y < \infty$, and

$$f_{x|y}(x \mid y) = \frac{\lambda^2 e^{-\lambda y}}{\lambda^2 y e^{-\lambda y}} = y^{-1} \tag{5.17}$$

for $0 < x \le y$. That is, X is uniformly distributed on $(0, y]$.

Note that

$$f_{xy}(x,y) = f_{y|x}(y \mid x) f_x(x) = f_{x|y}(x \mid y) f_y(y). \tag{5.18}$$

We can simulate data from this joint probability density by either one of the following two algorithms.

Algorithm 5.1

1) Draw X from the exponential density $f_x(x)$.
2) Draw Y from the exponential density $f_{y|x}(y \mid x)$ on the interval $[x, \infty)$.

Algorithm 5.2

1) Draw Y from the gamma density $f_y(y)$.
2) Draw X from the uniform density on the interval $[0, Y]$.

We will show later in this lecture that **Algorithm 5.2** states that conditional on (or given that) the sum of two exponential random variables which is a gamma random variable, the distribution of the first random variable is uniform.

Example 5.3 Dendritic Dynamics (Lee et al., PLoSB 4(2):e29 (2006)). To study the neuronal remodeling that occurs in the brain on a day-to-day basis, Elly Nedivi and colleagues used a multiphoton-based microscopy system for chronic in vivo imaging and reconstruction of entire neurons in the superficial layers of the rodent cerebral cortex. Over a period of months, they observed neurons extending and retracting existing branches, and in rare cases, budding new tips. Thirty-five of 259 non-pyramidal interneuron dendritic tips showed rearrangement and 0 of 124 pyramidal cell dendritic tips showed rearrangement. The objective is to analyze the probability of rearrangement for the two types of neurons.

Assume the propensity to change is a probability $p_i \in (0,1)$, where $i=1$ is the interneuron propensity and $i=2$ is the pyramidal neuron propensity. We will perform Bayesian analysis of this problem. Assume that p_i has a beta distribution with parameters α and β. This will represent our knowledge of p_i prior to the experiment and is called a prior distribution. Given p_i and X_i : Binomial (n_i, p_i) we have

$$f_{x_i|p_i}(x_i \mid p_i) = \binom{n_i}{x_i} p_i^{x_i}(1-p_i)^{1-x_i} \tag{5.19}$$

for $x_i = 0,1,...,n_i$ and $f(p_i)$ is the beta density

$$f(p_i) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{\alpha-1}(1-p_i)^{\beta-1}. \tag{5.20}$$

Let's compute $f(x_i)$ and $f(p_i \mid x_i)$. The joint distribution is the product of the pmf of a discrete random variable x_i and a continuous random variable p_i and is defined as

$$\begin{aligned}
f(x_i, p_i) &= f(p_i) f(x_i \mid p_i) \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p_i^{\alpha-1}(1-p_i)^{\beta-1} \binom{n_i}{x_i} p_i^{x_i}(1-p_i)^{n_i-x_i} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(n_i+1)}{\Gamma(x_i+1)\Gamma(n_i-x_i+1)} p_i^{\alpha+x_i-1}(1-p_i)^{n_i+\beta_i-x_i-1}
\end{aligned} \tag{5.21}$$

using the fact that if x_i is an integer $\Gamma(x_i) = (x_i-1)!$ by **Factoid 1** from the **Addendum to Lecture 3**

$$\binom{n_i}{x_i} = \frac{n_i!}{x_i!(n_i-x_i)!} = \frac{\Gamma(n_i+1)}{\Gamma(x_i+1)\Gamma(n_i-x_i+1)} \tag{5.22}$$

To compute $f(x_i)$ we consider

$$f(x_i) = \int_0^1 f(x_i, p_i) dp_i. \tag{5.23}$$

From **Lecture 3** and the **Addendum to Lecture 3**, we know that

$$\int_0^1 p_i^{\alpha+x_i-1}(1-p_i)^{n_i+\beta_i-x_i-1} dp_i = \frac{\Gamma(\alpha+x_i)\Gamma(n_i+\beta-x_i)}{\Gamma(\alpha+\beta+n_i)} \tag{5.24}$$

Hence,

$$f(x_i) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(n_i+1)}{\Gamma(x_i+1)\Gamma(n_i-x_i+1)} \frac{\Gamma(\alpha+x_i)\Gamma(n_i+\beta-x_i)}{\Gamma(\alpha+\beta+n_i)} \tag{5.25}$$

for $x_i = 0,1,2,...,n_i$. If we take $f(p_i)$ to be the beta probability density with $\alpha = \beta = 1$, then we have the uniform probability density on $[0,1]$. This means we are indifferent among all the values of p in $[0,1]$ prior to the start of the experiment and $f(x_i)$ simplifies to

$$f(x_i) = \frac{1}{n_i+1}. \tag{5.26}$$

Verify this.

For the Bayesian analysis we will be interested in $f(p_i \mid x_i)$, the most likely values of p_i given the data x_i for $i = 1,2.$ It is

$$\begin{aligned} f(p_i \mid x_i) &= \frac{f(p_i, x_i)}{f(x_i)} \\ &= \frac{\dfrac{\Gamma(\alpha+\beta)\Gamma(n_i+1)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(x_i+1)\Gamma(n_i-x_i+1)} p_i^{\alpha+x_i-1}(1-p_i)^{n_i+\beta-x_i-1}}{\dfrac{\Gamma(\alpha+\beta)\Gamma(n_i+1)\Gamma(\alpha+x_i)\Gamma(n_i+\beta-x_i)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(x_i+1)\Gamma(n_i-x_i+1)\Gamma(\alpha+\beta+n_i)}} \qquad (5.27) \\ &= \frac{\Gamma(\alpha+\beta+n_i)}{\Gamma(\alpha+x_i)\Gamma(n_i+\beta-x_i)} p_i^{\alpha+x_i-1}(1-p_i)^{n_i+\beta-x_i-1} \end{aligned}$$

which is a beta probability density with parameters $\alpha + x_i$ and $n_i + \beta - x_i$. We have now laid the ground work for an analysis we will be performing on these data in a few weeks.

III. Functions of Jointly Distributed Random Variables

The two functions of jointly distributed random variables X and Y that are most commonly considered are the sum and quotient

$$Z = X + Y$$
$$Z = \frac{Y}{X}. \tag{5.28}$$

We will consider only the sum. The quotient is discussed in Rice, pp. 98-99.

A. Sum of Two Random Variables

1. Discrete Random Variables

If X and Y are discrete random variables with joint pmf $p(X,Y)$ then $Z = X + Y$ is the set of all events $Z = z,\ X = x,\ Y = z - x.$ Hence,

$$p_z(z) = \sum_{x=-\infty}^{\infty} p(x, z-x) \tag{5.29}$$

or if $X + Y$ are independent

$$p_z(z) = \sum_{x=-\infty}^{\infty} p_x(x) p_y(z-x). \tag{5.30}$$

The sum is called the **convolution** of p_x and $p_{y.}$

Example 5.4. Suppose that the number of channel openings at two distant (independent) sites along the frog neuromuscular junction in a 500 msec interval are Poisson random variables with parameters λ_1 and λ_2. What is the probability mass function of the total number of channel openings in a 500 msec interval at the two sites?

We have $X : P(\lambda_1),\quad Y : P(\lambda_2)$ and let $Z = X + Y$ We want to consider the event $\{Z = X + Y = n\} = \bigcup_{k=0}^{n} \{X = k \cap Y = n - k\}$. Hence, we have

$$
\begin{aligned}
\Pr(X+Y=n) &= \sum_{k=0}^{n} \Pr(X=k \cap Y=n-k) \\
&= \sum_{k=0}^{n} \Pr(X=k)\Pr(Y=n-k) \\
&= \sum_{k=0}^{n} \frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!} \\
&= e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k} \\
&= \frac{(\lambda_1+\lambda_2)^n}{n!}e^{-(\lambda_1+\lambda_2)}.
\end{aligned}
\tag{5.31}
$$

We see that the sum of two independent Poisson random variables is again Poisson with parameters $\lambda_1+\lambda_2$. In general, if $Z=\sum_{j=1}^{J} X_j$ where $X_j \sim P(\lambda_j)$ then $Z \sim P(\sum_{j=1}^{J}\lambda_j)$. We will be able to establish this latter result rigorously once we develop the concept of the moment generating function in **Lecture 6**.

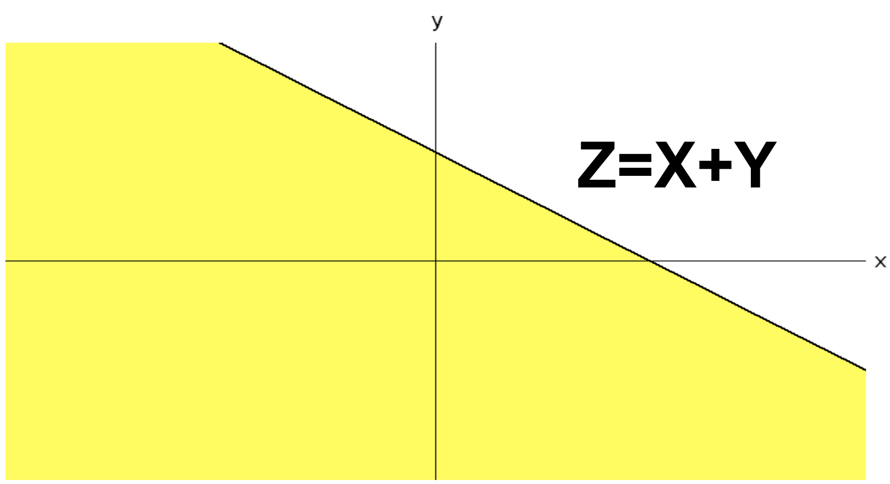

Figure 5.A. Set of X and Y values considered in the computation of a convolution of X+Y.

2. Continuous Random Variables

Now we consider in the continuous case $Z=X+Y\le z$ which is the region below the line $x+y=z$. Thus, if the joint pdf of X and Y is $f(x,y)$ we have

$$\begin{aligned} F_z(z) &= \Pr(X+Y \le z) \\ &= \iint_{R_z} f(X,Y)dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f(x,y)dydx \end{aligned} \tag{5.32}$$

Let $y = v - x,$ then $dy = dv$ and $y = z - x$ implies $v = y + x = z - x + x = z$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{z} f(x, v - x)dvdx. \tag{5.33}$$

If $\int_{-\infty}^{\infty} f(x, v-x)dx$ is continuous at $z,$ then on differentiating, we have

$$f_z(z) = \int_{-\infty}^{\infty} f(x, z-x)dx \tag{5.34}$$

and if X and Y are independent we obtain

$$f_z(z) = \int_{-\infty}^{\infty} f_x(x) f_y(z-x)dx \tag{5.35}$$

which is the **convolution** of f_x and $f_y.$

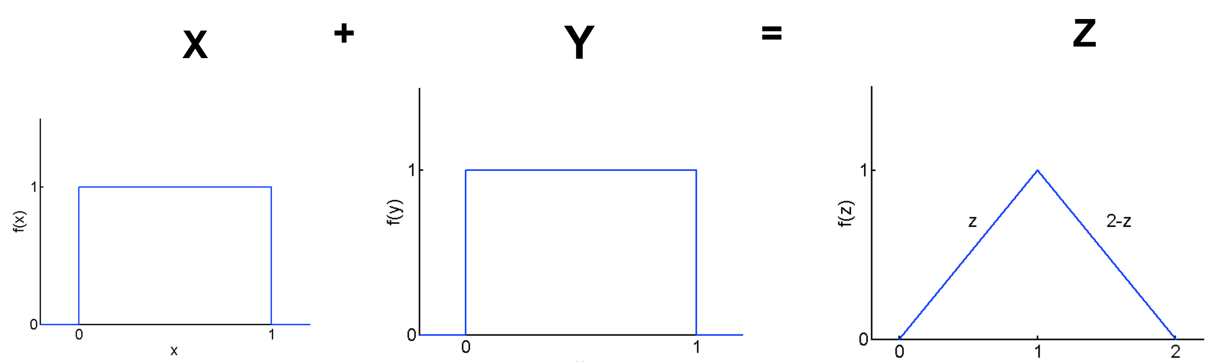

Figure 5B. Convolution of Two Uniform Random Variables Yields a Triangular Distribution.

Example 5.5. Suppose X and Y are independent random variables both uniformly distributed on $(0,1)$. Find $Z = X + Y$ (**Figure 5B**).

$$f_x(x)=\begin{cases}1 & 0<x<1\\ 0 & otherwise\end{cases}$$
$$f_y(y)=\begin{cases}1 & 0<y<1\\ 0 & otherwise\end{cases} \tag{5.36}$$

$z \in (0,2)$. To derive this density we must consider two cases:
Case i) $0<z\le 1$ then $x+y\le z$ and

$$f_z(z)=\int_0^1 f(x)f(z-x)dx=\int_0^z dx=z \tag{5.37}$$

Case ii) $1<z<2$ then $x+y=z$ such that $1<z<2$ then we must have $x\in(z-1,1)$ and

$$f_z(z)=\int_{z-1}^1 dx=x\Big|_{z-1}^1=1-(z-1)=2-z \tag{5.38}$$

$f(z)$ is the triangle probability density.

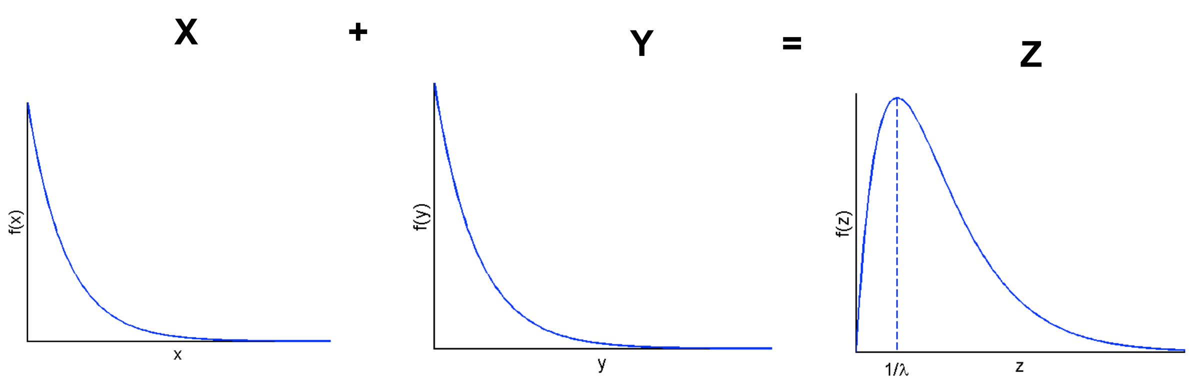

Figure 5C. Convolution of Two Exponential Random Variables Yields a Gamma Distribution.

Example 5.4 (continued). Suppose that the length of time that each of the two ion channels is open is an exponential random variable with parameters λ_1 for the first and λ_2 for the second. Because the channels are distant, we assume that they are independent. Find the probability density for the sum of the channel opening times if $\lambda_1=\lambda_2=\lambda$ (**Figure 5C**). Let $Z=X+Y$, then we have

$$f(x,y)=f_x(x)f_y(y)=\lambda^2 e^{-\lambda x}e^{-\lambda y} \tag{5.39}$$

for $x>0$ and $y>0$

$$f_z(z) = \int_{-\infty}^{\infty} f_x(x) f_y(z-x)dx = \int_0^z \lambda^2 e^{-\lambda x} e^{-\lambda(z-x)} dx$$
$$= \lambda^2 \int_0^z e^{-\lambda z} dx = \lambda^2 z e^{-\lambda z} \tag{5.40}$$

which is the probability density of a gamma random variable with $\alpha = 2$ and $\beta = \lambda$. Note that the convolution here, as in the case of the uniform random variables, “smooths” out the probability densities and makes values in the middle more likely (**Figure. 5C**). We saw the start of a similar phenomenon with the sum of two uniform random variables in **Example 5.5 (Figure 5B)** having a triangular density. This is a key observation for understanding the **Central Limit Theorem** in **Lecture 7**.

B. General Transformations of Two Random Variables
We state the general case as a Proposition.

Proposition 5.1. If X and Y are continuous random variables with joint pdf $f(x,y)$ suppose X and Y are mapped onto U and V by the transformation

$$u = g_1(x,y)$$
$$v = g_2(x,y) \tag{5.41}$$

and the transformation can be inverted to obtain

$$x = h_1(u,v)$$
$$y = h_2(u,v) \tag{5.42}$$

Assume that g_1 and g_2 have continuous partial derivatives, then the Jacobian is

$$J(u,v) = \begin{bmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{bmatrix}$$

and the absolute value of its determinant is

$$|J(u,v)| = \left| \frac{\partial h_1}{\partial u}\frac{\partial h_2}{\partial v} - \frac{\partial h_2}{\partial u}\frac{\partial h_1}{\partial v} \right| \tag{5.43}$$

then

$$f_{uv}(u,v) = f_{xy}(h_1(u,v), h_2(u,v)) \,|\, J(u,v) \,| \qquad (5.44)$$

for $u = g_1(x,y) \quad v = g_2(x,y)$ for some (x,y) and 0 elsewhere.

Notice that we used a form of **Proposition 5.1** to compute the normalizing constant for the beta distribution in the **Addendum to Lecture 3**.

Proof (Heuristic): The function g is a one-to-one mapping of a region in the (x,y) plane into a region in the (u,v) plane. It must be the case that the probability for any small region in the (x,y) plane must be equal to the probability of the corresponding small region in the (u,v) plane that it maps into under g. That is,

$$f_{uv}(u,v)d(u,v) = f_{xy}(x,y)d(x,y)$$

$$f_{uv}(u,v) = f_{xy}(x,y)\left|\frac{d(x,y)}{d(u,v)}\right|$$

$$f_{uv}(u,v) = f_{xy}(x,y) \,|\, J(u,v) \,|$$

$$f_{uv}(u,v) = f_{xy}(h_1(u,v), h_2(u,v)) \,|\, J(u,v) \,|$$

where, by analogy with the univariate change-of-variables in **Lecture 4**, we define $|J(u,v)|$ as the absolute value of the determinant of $J(u,v)$ to insure that $f_{uv}(u,v) \geq 0$.

Example 5.6. If X and Y are independent gamma random variables with parameters α and λ and β and λ respectively, find the joint pdf of $U = X + Y$ and $V = X(X+Y)^{-1}$.

The joint density of X and Y is given by

$$\begin{aligned} f_{x,y}(x,y) &= \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \frac{\lambda^{\beta}}{\Gamma(\beta)} y^{\beta-1} e^{-\lambda y} \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} y^{\beta-1} e^{-\lambda(x+y)} \end{aligned} \qquad (5.45)$$

Now we have $u = g_1(x,y) = x + y$, $v = g_2(x,y) = x(x+y)^{-1}$, $x = h_1(u,v) = uv$, and $y = h_2(u,v) = u(1-v)$. The Jacobian is

$$\begin{aligned} \frac{\partial h_1}{\partial u} &= v \qquad \frac{\partial h_1}{\partial v} = u \\ \frac{\partial h_2}{\partial u} &= 1-v \quad \frac{\partial h_2}{\partial v} = -u \end{aligned} \tag{5.46}$$

The absolute value of the determinant of the Jacobian is $|J(u,v)|=|-uv-u(1-v)|=u$. Thus,

$$\begin{aligned} f_{uv}(u,v) &= f_{xy}(uv,u(1-v))u \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)}(uv)^{\alpha-1}[u(1-v)]^{\beta-1}e^{-\lambda u}u \\ &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)}u^{\alpha+\beta-1}e^{-\lambda u}v^{\alpha-1}(1-v)^{\beta-1}. \end{aligned} \tag{5.47}$$

Multiplying and dividing by $\Gamma(\alpha+\beta)$ yields

$$\begin{aligned} f_{uv}(u,v) &= \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha+\beta)}u^{\alpha+\beta-1}e^{-\lambda u}\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}v^{\alpha-1}(1-v)^{\beta-1} \\ &= f_u(u)f_v(v) \end{aligned} \tag{5.48}$$

We see that $f_u(u)$ is the probability density of a gamma random variable with parameters $\alpha+\beta$ and λ, whereas $f_v(v)$ is the probability density of a beta random variable with parameters α and β. Therefore, the joint probability density $f_{uv}(u,v)$ is the product of a gamma and a beta density. By **Lecture 4**, because $f_{uv}(u,v)=f_u(u)f_v(v)$, we conclude that u and v are independent.

Notice that the arguments used in Eqs. 5.45 to 5.48 are essentially the arguments used to compute the normalizing constant for the beta probability density in the **Addendum to Lecture 3**.

Example 5.7. Let X and Y be independent standard Gaussian random variables. Find the joint density of $U=X$ and $V=X+Y$. We have

$$\begin{aligned} &g_1(x,y)=x \qquad g_2(x,y)=x+y \\ &x=h_1(u,v)=u \quad y=h_2(u,v)=v-u \end{aligned} \tag{5.49}$$

The Jacobian is

$$\frac{\partial h_1}{\partial u} = 1 \quad \frac{\partial h_1}{\partial v} = 0$$
$$\frac{\partial h_2}{\partial u} = -1 \quad \frac{\partial h_2}{\partial v} = 1$$

and the absolute value of its determinant is

$$|J(u,v)| = 1 \times 1 - 1 \times 0 = 1. \tag{5.50}$$

Therefore,

$$\begin{aligned} f_{uv}(u,v) &= f_{xy}(u, v-u)\,|J| = f_x(u) f_y(v-u)\,|J| \\ &= (2\pi)^{-1} \exp\{-\frac{1}{2}[u^2 + (v-u)^2]\} \times 1 \\ &= (2\pi)^{-1} \exp\{-\frac{1}{2}[u^2 + v^2 - 2uv + u^2]\} \\ &= (2\pi)^{-1} \exp\{-\frac{1}{2}[2u^2 + v^2 - 2uv]\}. \end{aligned} \tag{5.51}$$

Now matching up the terms on the right hand side of Eq. 5.51 with the terms in the joint density function of the bivariate Gaussian density in Eq. 4.32 yields

$$\begin{aligned} &\mu_u = \mu_v = 0 \\ &\sigma_u \sigma_v (1-\rho^2)^{\frac{1}{2}} = 1 \text{ and } \sigma_u^2 (1-\rho^2) = \tfrac{1}{2} \Rightarrow \sigma_v^2 = 2 \\ &\sigma_v^2 (1-\rho^2) = 1 \Rightarrow \rho^2 = \frac{1}{2} \text{ or } \rho = 2^{-\frac{1}{2}} \text{ and } \sigma_u = 1. \end{aligned} \tag{5.52}$$

Therefore, U and V have a bivariate Gaussian probability density with $\mu_u = \mu_v = 0, \sigma_u^2 = 1, \sigma_v^2 = 2$ and $\rho = 2^{-\frac{1}{2}}$.

Proposition 5.1 generalizes to n random variables (see Rice pp. 102-103).

IV. Summary

We have shown how to construct conditional pmf's and pdf's. In so doing, we have laid important groundwork for our statistical analyses. Similarly, understanding the convolution process that is necessary to compute the pmf or pdf of the sum of two independent random variables gives us key insight into how an important result like the **Central Limit Theorem** comes about.

Acknowledgments
I am grateful to Paymon Hosseini for making the figures, James Mutch for helpful editorial comments and Julie Scott for technical assistance.

References

Lee W, Huang H, Feng G, Sanes JR, Brown EN, So PT, Nedivi E. Dynamic remodeling of dendritic arbors in non-pyramidal neurons of adult visual cortex. *Public Library of Science – Biology*. 4(2)e29, 2006.

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