Physical Foundations of Physiology I: Pressure-Driven Flow 1.2

Learning Objectives

- Define intensive and extensive variables
- Define flow and flux
- Describe the driving principle for heat flow, electrical current, diffusive flow, and volume flow
- Explain what is meant by fluxes moving downhill
- Write a continuity equation and describe its meaning
- Explain why steady-state flux requires a linear gradient of T, Ψ, C, or P
- List four capacitances commonly encountered in physiology
- Define pressure and be able to convert pressure between atm, mmHg, and Pa
- Write Poiseuille's law, state its assumptions, and be able to calculate flow using it
- Write the Law of Laplace for cylindrical tubes and for spheres

FORCES PRODUCE FLOWS

EXTERNAL AND INTERNAL MOVEMENT IS A HALLMARK OF HUMAN LIFE

For humans in their natural environment, movement is essential for survival. This movement refers to translation of the body from one location to another, and movement of the limbs relative to one another. In addition to this movement of body parts with respect to the external world, movement of materials within the body is also essential. Most important among these internal movements are the movement of the blood, movement of the air in and out of the lungs, movement of food and fecal material along the gastrointestinal tract, and movement of the urine from its formation to elimination. In addition, the body transports materials across barriers such as the gastrointestinal tract lining, lungs, and kidney tubule. Transport also occurs within cells. All of these movements require the continued application of force to overcome inertia and friction.

TRANSPORT OF MATERIAL IS DESCRIBED AS A FLOW OR A FLUX

The transport of material is quantitatively expressed as a **flow**, which we will symbolize by the variable Q. The flow can be expressed as

- volume of material or fluid transported per unit time;
- mass of material transported per unit time;
- number of particles or moles transported per unit time;
- number of ions or unit charges transported per unit time (electrical current).

FLOW DEPENDS ON THE AREA; FLUX IS FLOW PER UNIT AREA

The total flow of volume or solute is an **extensive** variable: the flow depends on the extent or the amount of the system that gives rise to the flow. In the case of two compartments separated by a membrane, doubling the area or extent of the membrane would produce twice as much flow between the two compartments. Dividing the flows by the area normalizes the flows. The normalized flow is the flux, and the flux is an **intensive** variable whose value is independent of the extent of the system. Flux is *defined* as

$$J_V = \frac{Q_V}{A} \qquad J_S = \frac{Q_S}{A} \tag{1.2.1}$$

where Q_V is the volume flow, J_V is the volume flux, A is the cross-sectional area through which flow occurs, oriented at right angles to the direction of flow, Q_S is the amount of material (solute) transported per unit time, J_S is the solute flux, and A is the area. The units of flux are amount or volume or mass or charge per unit time per unit area.

Strictly speaking, fluxes and flows are vectors, consisting of the magnitude of the flux or flow and its direction. Unless otherwise noted, we will consider flux or flow only in one direction and therefore we will suppress the vector nature of flux and flow.

DOI: http://dx.doi.org/10.1016/B978-0-12-800883-6.00002-1

FLUX DEPENDS LINEARLY ON ITS CONJUGATE FORCE

For a variety of forces and fluxes, the flux that results from a net force varies linearly with the force:

$$J_x = LF_x \tag{1.2.2}$$

where J_x is the flux of something, L is a phenomenological coefficient, and F_x is the net force that drives the flux. This generic equation holds for a variety of kinds of fluxes. The flux of heat energy, electrical flux (the current density), diffusion of solute and pressure-driven flow all obey this general phenomenological law. In each of these cases, the net force is proportional to the **gradient** of an intensive variable. Strictly speaking, the gradient is a vector quantity, but we use it here to denote the slope of these intensive variables along one dimension:

$$\begin{aligned}
&\text{Fourier's law of heat conduction:} && J_H = -\lambda \frac{dT}{dx}\\
&\text{Ohm's current law:} && J_e = -\sigma \frac{d\psi}{dx}\\
&\text{Fick's first law of diffusion:} && J_S = -D \frac{dC}{dx}\\
&\text{Pressure-driven flow:} && J_V = -L_P \frac{dP}{dx}
\end{aligned} \tag{1.2.3}$$

where J_H is the flux of heat energy, dT/dx is the temperature gradient, λ is the coefficient of thermal conductivity, J_e is the electrical current flux, $d\psi/dx$ is the voltage gradient, σ is the electrical conductivity, J_S is the solute flux, dC/dx is the concentration gradient, D is the diffusion coefficient, J_V is the volume flux, dP/dx is the pressure gradient, and L_P is the hydraulic conductivity. All of these phenomenological equations find application in physiological systems. They are all analogues of Ohm's law.

These equations are true only if the only driving force is the one specified. For example, diffusion of electrolytes, charged solutes, is influenced by electric fields. If a voltage gradient is also present along with a concentration gradient, Fick's first law of diffusion would need to be modified to reflect that influence. A pressure difference that produces a volume flow in the presence of a diffusion gradient will also modify the flux of solute. In general, flows produced by multiple flow processes are not independent. If there are two forces driving flows, we write

$$\begin{aligned} J_1 &= L_{11}F_1 + L_{12}F_2\\ J_2 &= L_{21}F_1 + L_{22}F_2 \end{aligned} \tag{1.2.4}$$

where L_{11} is the coefficient relating flux 1 to its primary driving force 1 and L_{22} relates flux 2 to its primary driving force 2, and L_{12} and L_{21} are the coupling coefficients that describe how secondary forces affect the flows. An example of this is a bimetallic junction. When two unlike metals are joined together, passing a current through the junction causes it to either heat up or cool, and this is called the *Peltier effect*. The coupling coefficient implies that if you heat up or cool the junction, a current will flow. This is the basis of the *thermocouple*. In this case, the two fluxes are heat and current and the two forces are temperature gradient and voltage gradient. Because of a principle called microscopic reversibility, it turns out that the cross-coupling coefficients are equal: $L_{12} = L_{21}$. This is called **Onsager reciprocity**, in honor of Lars Onsager (1903–1976) who earned the Nobel Prize in Chemistry in 1968 for this discovery.

FLUX MOVES DOWNHILL

The relations in Eqn [1.2.3] describe fluxes in one dimension. Both fluxes and gradients are actually vectors, but we consider a single direction here for simplicity. Consider Fick's law of diffusion for solutes. If the gradient of concentration is constant, we may write

$$J_S = -D\frac{(C_1 - C_2)}{(x_1 - x_2)} \tag{1.2.5}$$

for two points (C_1, x_1) and (C_2, x_2). If $C_1 > C_2$ and $x_1 < x_2$, the slope is negative and the flux is positive. If $C_1 < C_2$ and $x_1 < x_2$, then the slope is positive and the flux is negative. Thus, the flux always goes from regions of high concentration to regions of low concentration (see Figure 1.2.1). This is true for all the intensive variables for the fluxes in Eqn [1.2.3]. These fluxes always move downhill, unless acted upon by additional forces.

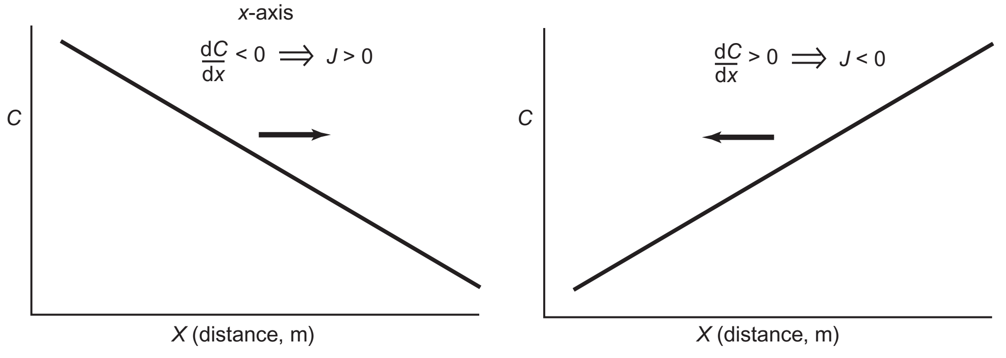

FIGURE 1.2.1 Flux moves downhill. Consider one-dimensional flux, with positive flux defined as in the direction of the *x*-axis. In this case, we consider diffusion that is driven only by a concentration gradient. If the gradient is negative (higher concentrations at lower values of *x*), then by Eqn [1.2.3], the flux is positive and directed to the right (middle panel). If the gradient is positive, then the flux is negative and directed to the left. In each case, the flux of solute is from the region of high concentration toward the region of lower concentration.

CONSERVATION OF MATTER OR ENERGY LEADS TO THE *CONTINUITY EQUATION*

Here we consider that a concentration gradient exists and produces a solute flux as a consequence. We consider a cylindrical tube as shown in Figure 1.2.2, having a cross-sectional area A, which is intersected at right angles by planes at $x = x$ and $x = x + \Delta x$. so that the tube is cut into three compartments, the left, middle, and right compartments.

The concentration may vary with time and distance. We define the concentration in any volume element as

[1.2.6] $$C(x,t) = \frac{N(x,t)}{V}$$

where $N(x,\ t)$ is the number of solute particles in the volume element and V is the volume element. We define $J(x)$ as the *net* number of solute particles crossing the plane at $x = x$ per unit time per unit area, with positive being directed along the x-axis, to the right. The number of particles entering the middle compartment from the left in time Δt is $AJ(x)\Delta t$ and the number leaving the middle compartment by crossing the plane at $x = x + \Delta x$ is $AJ(x + \Delta x)\Delta t$. Here the parenthesis means "function of" and not multiplication. If there is no chemical transformation of the solute particles, their number is conserved and we can write

[1.2.7] $$\begin{aligned}\Delta N &= AJ(x)\Delta t - AJ(x+\Delta x)\Delta t\\ \frac{\Delta N}{\Delta t} &= AJ(x)t - AJ(x+\Delta x)\end{aligned}$$

Dividing by the volume element $V = A\Delta x$ and rearranging, we have

[1.2.8] $$\begin{aligned}\frac{1}{A\Delta x}\frac{\Delta N}{\Delta t} &= \frac{AJ(x) - AJ(x+\Delta x)}{A\Delta x}\\ \frac{\Delta \frac{N}{V}}{\Delta t} &= -\frac{AJ(x+\Delta x) - AJ(x)}{A\Delta x}\end{aligned}$$

In the limit of $\Delta t \to 0$ and $\Delta x \to 0$, this becomes

[1.2.9] $$\frac{\partial C(x,t)}{\partial t} = -\frac{\partial J(x)}{\partial x}$$

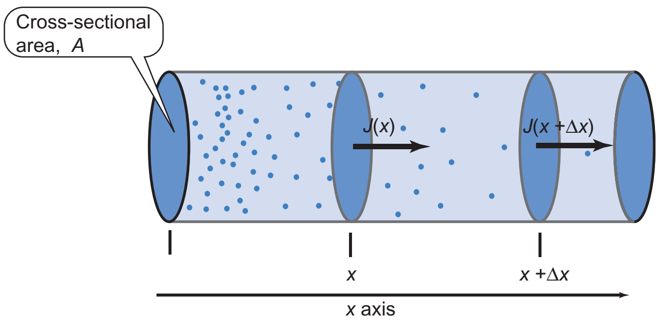

FIGURE 1.2.2 Fluxes as a function of distance in the presence of a concentration gradient.

This equation is called the **continuity equation**. What this equation says is that if the flux of solute is not the same everywhere, then the amount of solute must be building up or becoming depleted somewhere, and this buildup or depletion changes the concentration of solute. It is a straightforward consequence of the conservation of material. This equation is not true if the diffusing chemical undergoes chemical transformation. In this case, it is not conserved.

Similar continuity equations can be written for the flux of heat energy, charge, and volume. Their form is given in Eqn [1.2.10]:

[1.2.10] $$\begin{aligned}&\text{Heat:}\quad \rho C_P\frac{\partial T}{\partial t} = -\frac{\partial J_H}{\partial x}\\ &\text{Charge:}\quad \frac{C}{V}\frac{\partial \psi}{\partial t} = -\frac{\partial J_e}{\partial x}\\ &\text{Solute:}\quad \frac{\partial C_S}{\partial t} = -\frac{\partial J_S}{\partial x}\\ &\text{Volume:}\quad \frac{C}{V}\frac{\partial P}{\partial t} = -\frac{\partial J_V}{\partial x}\end{aligned}$$

where ρ is the density of matter through which heat flows and C_p is its specific heat capacity. In the next equation, C is the electrical capacitance, V is the volume. Next, C_s is the concentration. In the last line C is the compliance, and V is the volume. Here we have the unfortunate situation in which single variables denote different quantities: C stands for electrical capacitance (in farads), concentration (usually in molar) and compliance ($=\Delta V/\Delta P$). Generally the meaning of the variable is clear from its context.

STEADY-STATE FLOWS REQUIRE LINEAR GRADIENTS

In homeostasis, there is a steady supply of nutrients and removal of wastes and a steady withdrawal of nutrients and a steady production of wastes by the tissues. This steady state in which all flows are constant is more easily amenable to mathematical analysis. What "steady state" means is that each of the variables on the left-hand side of Eqn [1.2.10] is zero because at the steady state there are no changes in temperature, charge, concentration, or pressure with time:

[1.2.11] $$\begin{aligned}&\text{Heat:}\quad \rho C_P\frac{\partial T}{\partial t} = -\frac{\partial J_H}{\partial x} = 0\\ &\text{Charge:}\quad \frac{C}{V}\frac{\partial \psi}{\partial t} = -\frac{\partial J_e}{\partial x} = 0\\ &\text{Solute:}\quad \frac{\partial C_S}{\partial t} = -\frac{\partial J_S}{\partial x} = 0\\ &\text{Volume:}\quad \frac{C}{V}\frac{\partial P}{\partial t} = -\frac{\partial J_V}{\partial x} = 0\end{aligned}$$

Substituting in from Eqn [1.2.3] for J_H, J_e, J_S, and J_V, we have

$$\frac{\partial^2 T}{\partial x^2} = 0$$
$$\frac{\partial^2 \psi}{\partial x^2} = 0$$
[1.2.12]
$$\frac{\partial^2 C}{\partial x^2} = 0$$
$$\frac{\partial^2 P}{\partial x^2} = 0$$

This condition is met only if the gradient of T, ψ, C, or P is constant; thus, the slope of T, ψ, C, and P at steady state is constant, and each of these intensive variables varies linearly with distance.

HEAT, CHARGE, SOLUTE, AND VOLUME CAN BE STORED: ANALOGUES OF *CAPACITANCE*

The steady state is often approximated in the body but rarely achieved. At rest heat production balances heat exhausted to the environment. When we begin exercising, heat production rises rapidly and the temperature of the body rises accordingly until, once again, heat production matches heat transfer to the environment, achieved by using other forces besides the conduction described in Fourier's law. This new steady state of temperature during exercise is achieved at different operating conditions than at rest. In another example, transport of blood through the cardiovascular system is pulsatile, because the pressure that drives transport comes from the heart, and the heart produces force rhythmically. Each of the main four variables we have been discussing, heat, charge, amount of chemicals, and volume, can be temporarily stored or depleted.

Electrical charge can be stored in capacitors. The constitutive relation between charge, voltage, and capacitance is given as

[1.2.13] $$Q = CV$$

where Q here stands for charge, C is the capacitance, and V is the voltage. Here we are victims of the use of the same variables to denote entirely different quantities. We will use Q most often to signify a flow, but here it signifies charge, in coulombs. In physiology, we often use C to denote concentration, but here it means capacitance, in farads ($=CV^{-1}$); in physiology, V usually signifies volume, but here it means electrical potential, in volts. Electrical capacitance is an important concept for physiologists as well, because membrane potential derives from a separation of electrical charges across the membrane, and the membrane itself acts like a tiny capacitor with two conducting plates, separated by a dielectric. We will discuss this further in the sections on membrane potential, action potential, and the cable properties of nerves (Chapters 3.1–3.3). The other relationships completely analogous to the relation between charge, capacitance, and voltage, are

$$H = C_P MT\text{: Heat energy} = \text{heat capacity} \times \text{mass} \times \text{temperature}$$
$$M = VC\text{: Amount} = \text{volume} \times \text{concentration}$$
$$V = CP\text{: Volume} = \text{compliance} \times \text{pressure}$$
[1.2.14]

where the capacitance-like elements include electrical capacitance (C in Eqn [1.2.13]), thermal mass (C_pM, the specific heat capacity times the mass), volume (V), and compliance (C). Note again the multiplicity of uses of a single notation. C variously stands for capacitance, heat capacity, concentration, or compliance.

The capacitances are all expressed as the ratio of an extensive variable and an intensive variable and are all themselves extensive variables. Table 1.2.1 summarizes the four kinds of capacitances.

PRESSURE DRIVES FLUID FLOW

In the case of fluid or air flow, pressure differences drive the flow. The SI unit of pressure is the pascal, Pa, equal to 1 N m^{-2}. However, physiologists still use other units, notably the atmosphere and mmHg. The atmospheric pressure is the weight of a column of air equal to the height of the atmosphere in the earth's gravitational field per unit area of the earth's surface. The actual pressure in the atmosphere decreases as you ascend, but the unit of 1 atm is defined for a standard condition of the air and standard altitude at sea level. The conversion between atmospheres and mmHg is an observed phenomenon. Atmospheric pressure can be measured in units of mmHg as described in Figure 1.2.3.

Figure 1.2.3 illustrates that atmospheric pressure supports a column of 760 mmHg high. The pressure of this column of Hg is equal to atmospheric pressure, and the pressure of the column is simply its weight divided by its area. The weight is the force of gravity acting on the column, and is given as

[1.2.15]
$$\begin{aligned} W = F &= mg \\ &= \rho V g \\ &= \rho A h g \end{aligned}$$

TABLE 1.2.1 Four Kinds of Capacitances

Capacitance	Expression	Units	Application
Electrical **capacitance**, C	Q/V	Farads = coulombs/volt	Nerve conduction, membrane potential
Thermal capacitance, C_pM	H/T	JK^{-1} = joules/temperature	Heat production/loss temperature regulation
Chemical capacitance, V, **volume**	M/C	Volume = moles/molarity, L	Metabolism, filtration
Mechanical capacitance, C, **compliance**	V/P	Compliance = volume/pressure, m^3 Pa^{-1}	Blood pressure, breathing

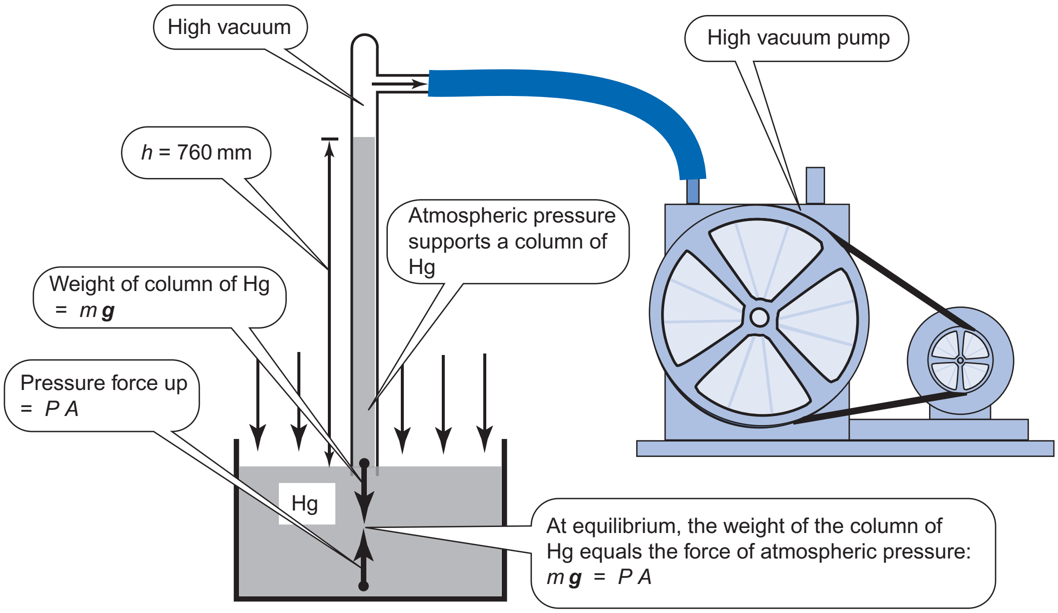

FIGURE 1.2.3 Measurement of atmospheric pressure. A closed vertical tube is connected to a high vacuum pump and evacuated air. When inverted into a dish of mercury, the atmospheric pressure forces mercury up the tube until mechanical equilibrium is achieved when the weight of the column of mercury exerts a pressure equal to the atmospheric pressure. At sea level in dry air, 1 atm will support a column 760 mmHg high.

TABLE 1.2.2 Conversion Between Pressure Units

atm	mmHg	Pa
1	760	1.013×10^5
0.00132	1	133.29
9.87×10^{-6}	0.00750	1

The pressure is just the force per unit area. Dividing Eqn [1.2.15] by the area, we get

$$P = \frac{F}{A} = \rho g h \tag{1.2.16}$$

Thus, the height of the column of mercury in equilibrium with the atmospheric pressure is independent of its area. We need to specify only the height of the column of mercury. Thus, at sea level, the atmospheric pressure supports a column of 760 mmHg high and we say that 760 mmHg = 1 atm.

The value of atmospheric pressure in pascals = N m^{-2} can be calculated from 760 mmHg by using the density of Hg (13.59 g cm^{-3}) and the acceleration due to gravity (9.81 m s^2). Inserting these values into Eqn [1.2.16], we get

$$\begin{aligned} P &= 13.59\ \mathrm{g\ cm^{-3}} \times 9.81\ \mathrm{m\ s^{-2}} \times 0.76\ \mathrm{m} \\ &= 13.59\ \mathrm{g\ cm^{-3}} \times (100\ \mathrm{cm\ m^{-1}})^3 \times 10^{-3}\ \mathrm{kg\ g^{-1}} \\ &\quad \times 9.81\ \mathrm{m\ s^{-2}} \times 0.76\ \mathrm{m} \\ &= 13.59 \times 10^3\ \mathrm{kg\ m^{-3}} \times 9.81\ \mathrm{m\ s^{-2}} \times 0.76\ \mathrm{m} \\ 1\ \mathrm{atm} &= 101.3 \times 10^3\ \mathrm{kg\ m\ s^{-2}\ m^{-2}} = 1.013 \\ &\quad \times 10^5\ \mathrm{N\ m^{-2}} = 1.013 \times 10^5\ \mathbf{Pa} \end{aligned}$$

We can therefore complete a conversion table for pressure units (Table 1.2.2).

POISEUILLE'S LAW GOVERNS STEADY-STATE LAMINAR FLOW IN NARROW TUBES

In 1835, Jean Leonard Marie Poiseuille experimentally established the relationship between flow through narrow pipes and the pressure that drives the flow. The relationship is

$$Q_V = \frac{\pi a^4}{8\eta}\left(\frac{\Delta P}{\Delta x}\right) \tag{1.2.17}$$

where Q_V is the flow, in units of volume per unit time, π is the geometric ratio, a is the radius of the pipe, η is the viscosity, ΔP is pressure difference between the beginning and end of the pipe, and Δx is the length of the pipe. This equation describes the relationship between flow and pressure difference only for laminar flow. Laminar flow is steady, streamlined flow, and it is distinguished from turbulent or chaotic flow. This equation is often applied to problems in physiology even though the conditions for its valid application are missing. Its application requires us to understand **viscosity**.

Consider two parallel plates separated by a fluid, as shown in Figure 1.2.4. The top plate can be moved at a constant velocity relative to the stationary bottom plate only if the plate is subjected to a force that continuously overcomes the frictional resistance on the plate caused by its contact with the adjacent fluid.

The viscosity is the resistance of a fluid to shear forces. It is defined by the equation

$$\frac{F}{A} = \eta \frac{dv}{dy} \tag{1.2.18}$$

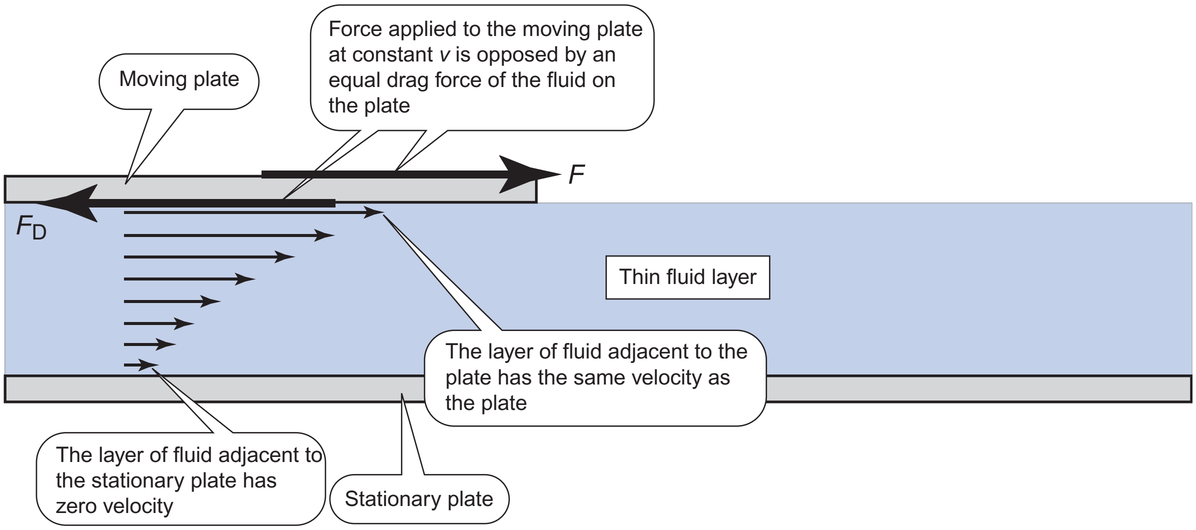

FIGURE 1.2.4 Definition of viscosity. Two plates are separated by a fluid. The top plate moves with constant velocity, v, with respect to the stationary bottom plate. The fluid adheres to the plates and a thin layer of fluid immediately adjacent to the plates has the same velocity as the plates. This results in a velocity profile in the fluid. The steepness of this velocity profile, dv/dy, is the velocity gradient.

where F is the shear force, A is the area, v is the velocity, and y is the dimension perpendicular to the plate. The ratio F/A is called the **shear stress** and the quantity dv/dy is called the **velocity gradient**. F/A in this equation has the units of pressure, $\text{Pa} = \text{N m}^{-2}$ and dv/dy has the units of $\text{m s}^{-1}\ \text{m}^{-1} = \text{s}^{-1}$, so the units of η in SI are Pa s. Thus, the viscosity is the ratio of the shear stress, F/A, dvided by the velocity gradient, dv/dy. In older texts viscosity is sometimes given in units of poise = 1 dyne cm^{-2} s. These can be converted to Pa s by using the definition of $\text{Pa} = 1\ \text{N m}^{-2}$, $1\ \text{N} = 1\ \text{kg m s}^{-2}$ and 1 dyne = $1\ \text{g cm s}^{-2}$: Thus, $1\ \text{N} = 10^5$ dyne.

$$
\begin{aligned}
&1\ \text{Pas} = 1\ \text{N m}^{-2}\ \text{s} = 1\ \text{kg m s}^{-2}\ \text{m}^{-2}\ \text{s} = 1\ \text{kg}\\
&\times 10^3\ \text{g kg}^{-1}\ \text{m} \times 100\ \text{cm m}^{-1} \times \text{s}^{-2} \times \text{m}^{-2}\\
&\times (0.01\ \text{cm m}^{-1})^2\ \text{s} = 10^3\ \text{g} \times 100\ \text{cm} \times \text{s}^{-2}\\
&\times 10^{-4}\ \text{cm}^{-2}\ \text{s} = 10\ \text{g cm s}^{-2} - \text{cm}^{-2}\ \text{s} =\\
&10\ \text{dyne cm}^{-2}\ \text{s} = 10\ \text{poise}\\
&\text{So } 1\ \text{Pa s} = 10\ \text{poise}
\end{aligned}
$$

EXAMPLE 1.2.1 Ultrafiltration in the Kidney

The kidney has a structure called the glomerulus that consists of combined layers of cells and extracellular matrix—bundles of fibers in the extracellular space—that together form an ultrafilter (see Chapter 6.2 for further description). It is called an ultrafilter because the combined layers retain proteins while letting most small solutes pass into the ultrafiltrate. We model the membrane here as a flat membrane that is pierced by many identical right cylindrical holes, or pores. Assume that the radius of the pores is 3.5 nm and the pore length is 50 nm. The viscosity of the fluid is taken to be the same as plasma, 0.02 poise. The aggregate area of the pores makes up 5% of the total surface area of the membrane. The total pressure on the input side, the side of the blood, averages 60 mmHg and on the ultrafiltrate side the total pressure averages 45 mmHg (see Chapter 6.2 for a discussion of the origin of these pressures). The total available area of the membrane is 1.5 m^2. What is the filtration rate in $\text{cm}^3\ \text{min}^{-1}$?

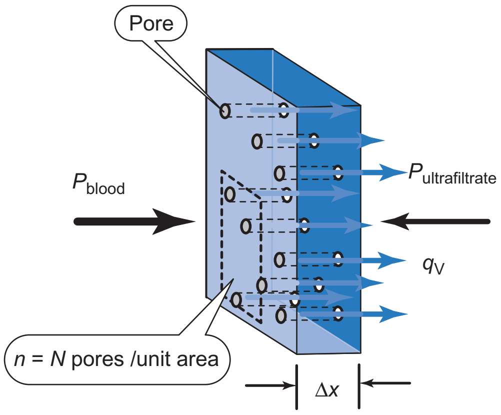

FIGURE 1.2.5 Model of the kidney ultrafilter.

The situation is depicted schematically in Figure 1.2.5. We use Poiseuille's equation here. The total flow, Q_V, is the sum of the flow through all of the pores:

$$Q_v = Nq_v = N\pi a^4/8\eta\,(\Delta P/\Delta x)$$

where N is the number of pores and q_V denotes the flow through a single pore. Here a is the radius, $a = 3.5 \times 10^{-9}$ m, η is the viscosity ($\eta = 0.02$ poise $\times$ 1 Pa s/10 poise = 0.002 Pa s), ΔP is the pressure difference = 15 mmHg $\times$ 133.3 Pa mmHg^{-1} = 1999.5 Pa, and $\Delta x = 60 \times 10^{-9}$ m. Now that all the units are compatible, we plug them into the equation and get:

$$q_V = \pi \times (3.5 \times 10^{-9}\ \text{m})^4 / 8 \times 2 \times 10^{-3}\ \text{Pa s} \times (1999.5\ \text{Pa}/60 \times 10^{-9}\ \text{m})$$
$$= 471.44 \times 10^{-36}\ \text{m}^4 / 16 \times 10^{-3}\ \text{Pa s} \times 33.33 \times 10^9\ \text{Pa} \times \text{m}^{-1}$$
$$= 982.1 \times 10^{-24}\ \text{m}^3 \times \text{s}^{-1} \times (100\ \text{cm} \times \text{m}^{-1})^3 \times 60\ \text{s} \times \text{min}^{-1}$$
$$= 5.89 \times 10^{-14}\ \text{cm}^3 \times \text{min}^{-1}$$

This is the flow through a single pore. We need to know how many of them are there. If their aggregate area is 5% of the total, then the number of pores can be calculated from

$$N \times \pi a^2 = 0.05 \times 1.5\ \text{m}^2$$

Knowing that $a = 3.5 \times 10^{-9}$ m, we solve for N:

$$N = 0.075\ \text{m}^2 / \pi \times (3.5 \times 10^{-9}\ \text{m})^2 = 1.95 \times 10^{15}$$

which is a lot of pores! Multiplying $N \times q_V$, we get

$$Q_V = N \times q_V = 1.9 \times 1015\ \text{pores} \times 5.89 \times 10^{-14}\ \text{cm}^3\ \text{min}^{-1}\ \text{pore}^{-1}$$
$$= \mathbf{111.9\ cm^3\ min^{-1}}$$

This is a reasonable approximation to the filtration rate in an adult human.

THE *LAW OF LAPLACE* RELATES PRESSURE TO TENSION IN HOLLOW ORGANS

The blood vessels maintain a pressure difference across their walls. These vessels approximate hollow cylinders. The gallbladder, urinary bladder, heart, and lung alveoli also maintain a pressure difference, and approximate hollow spheres. The Law of Laplace can be derived for these ideal geometries by considering mechanical equilibrium. Consider first a circular cylinder of radius r subjected to an external pressure P_o and internal pressure P_i, as shown in Figure 1.2.6.

In Figure 1.2.6, the internal pressure acting on the upper half of the cylinder by the lower half cylinder produces a force on the upper half that is given as $P_i \times 2r \times L$. The external pressure acting over the surface of the upper half of the cylinder produces a force that is equal to $P_o \times 2r \times L$, as can be deduced from the situation where $P_o = P_i$ in a mechanically stable fluid with no walls. The balance of the internal and external pressures produces a net force of $\Delta P \times 2r \times L$ where $\Delta P = P_i - P_o$. This is balanced by the force of the walls of the lower half acting on the upper half. These forces are the wall tension, in units of $\text{N} \times \text{m}^{-1}$, acting along the length of the wall. These forces are directed downward, as shown. For mechanical equilibrium to occur

[1.2.19] $$F_{net} = 0 = T2L = \Delta P 2rL$$

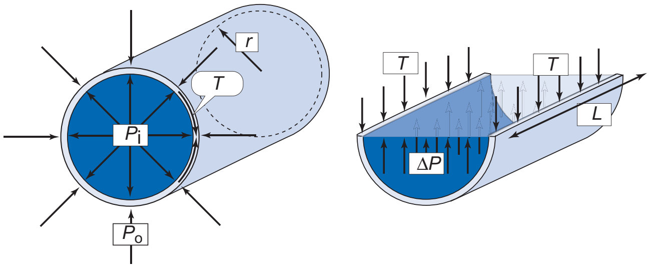

FIGURE 1.2.6 Right circular cylinder of radius r and length L is subjected to a transmural pressure difference. For mechanical equilibrium, the net force on any part of the cylinder must be zero. We split the cylinder in half down its axis. The sum of forces on the upper half must sum to zero.

Rearrangement of this equation gives the Law of Laplace for a cylinder:

[1.2.20] $$\Delta P = \frac{T}{r}$$

An exactly analogous argument can be made for a sphere. In this case, the pressure is distributed over the cross-sectional area of the hemisphere, and the wall tension is distributed around the circumference. The result gives

[1.2.21] $$F_{net} = 0 = T2\pi r = \Delta P \pi r^2$$

which is easily rearranged to

[1.2.22] $$\Delta P = \frac{2T}{r}$$

SUMMARY

Flow consists of the movement of something from one place to another in the body. We identify four classes of flow of major importance in physiology: flow of heat energy, electric current, solute, and volume (gas or body fluids). Flow is an extensive variable, depending on the extent of the system. Flux is the flow divided by the area through which the flow occurs. Fluxes are driven by forces. Heat flux is driven by a temperature gradient, electrical current by a potential gradient, solute flux by a concentration gradient, and volume flux by a pressure gradient. Gradients are vector quantities. Here we consider one-dimensional flow and flux and suppress the vector nature of these gradients. In each of these cases, flux occurs "downhill," meaning from regions of high temperature, potential, concentration or pressure, to regions of low temperature, potential, concentration, or pressure. Each flux has the form

$$J_x = -K \frac{d\psi_x}{dx}$$

The continuity equation for each of these fluxes states that a gradient of flux produces a time dependence of the driving force. If the flux is not the same everywhere, then temperature, voltage, concentration, or pressure varies with time. The continuity equation is a consequence of conservation of energy, electric charge, chemical species, or volume.

Steady-state flows require linear gradients of temperature, potential, concentration, or pressure.

Heat energy, electric charge, concentration, and volume can all be stored in the body. The ability to store these things is quantified by capacitances. These include thermal capacitance, electrical capacitance, chemical capacitance, and mechanical capacitance. The thermal capacitance is MC_p, where C_p is the specific heat capacity; the units of MC_p are in J K^{-1}; electrical capacitance is Q/V, in F; chemical capacitance is V, in L; mechanical capacitance is compliance, in L Pa^{-1}.

Pressure drives flow. Pressure is measured in pascals = 1 N m^{-2}. This unit is equal to 9.87×10^{-6} atm or 0.0075 mmHg. Steady-state pressure-driven flow through narrow pipes is described by Poiseuille's law:

$$Q_V = \frac{\pi a^4}{8\eta}\left(\frac{\Delta P}{\Delta x}\right)$$

where a is the radius of the pipe, η is the viscosity, in Pa s, π is the geometric ratio of circumference to diameter of a circle, ΔP is the pressure difference in the pipe, and Δx is the length of the pipe. Viscosity is the resistance of a fluid to shear forces and is defined mathematically as the ratio of the shear stress to the velocity gradient.

A transmural pressure within hollow organs or tubes requires tension in the walls of these organs. So far as these organs approximate thin-walled cylinders or spheres, the wall tension obeys the Law of Laplace:

$$\text{Cylinder:} \quad \Delta P = \frac{T}{r}$$

$$\text{Sphere:} \quad \Delta P = \frac{2T}{r}$$

where ΔP is the transmural pressure difference, T is the wall tension, in N m^{-1}, and r is the radius of the cylinder or sphere.

REVIEW QUESTIONS

1. Is density an intensive or extensive variable? Is temperature intensive or extensive?
2. What drives current flow? What drives solute flow? If solute is charged, would its movement make a current?
3. What does it mean to say that fluxes "move downhill"?
4. What does capacitance mean for charge? Solute? Heat? Volume?
5. How is Ohm's law like Fick's law of diffusion?
6. What are the units of viscosity?
7. What are the assumptions in the derivation of Poiseuille's law? Are these reasonable?
8. What is the relationship between velocity and flux in fluid flow?
9. What is the relationship between capacitance, charge, and voltage?
10. What is the relationship between volume, amount, and concentration?
11. What is the relationship between compliance, volume, and pressure?
12. What is the relationship between heat capacity, heat energy, and temperature?
13. What is the relationship between wall tension and pressure in a sphere? In a cylinder?

APPENDIX 1.2.A1 DERIVATION OF POISEUILLE'S LAW

POISEUILLE'S LAW DESCRIBES PRESSURE-DRIVEN FLOW THROUGH A CYLINDRICAL PIPE

Consider two fluid compartments which are joined by a right cylindrical pipe, as shown below, of area A. Since pressure is defined as a force per unit area, the total force acting on the fluid in the pipe on the left-hand surface is just $P_L A$, and the total force acting on the fluid in the pipe on the right-hand side is $P_R A$. If P_L and P_R are not equal, then the fluid within the pipe will be subject to a net force and therefore this volume of fluid will be accelerated. The result will be a movement of fluid in the direction of the net force. What we wish to establish is the quantitative relationship between the resulting flow and the pressure difference which drives this flow.

SHEAR STRESS IS THE VISCOSITY TIMES THE VELOCITY GRADIENT

The movement of fluid through the pipe shown in Figure 1.2.A1.1 encounters resistance along the cylindrical surfaces of the pipe. This resistance is key to deriving an expression for the flow as a function of pressure. It is due to the viscosity of the fluid, as shown in Figure 1.2.4. We consider that the flow of fluid occurs in layers, or laminae, and that each layer exerts a viscous drag on the layer immediately above and below. The fluid remains in contact with the walls on either side of the tube, and so has a velocity $v = 0$ at the walls.

In order to achieve a constant velocity, we need to apply a continual force, F, in order to overcome the viscous drag of the fluid. The **shear stress** is the force exerted by one lamina on an adjacent one, and is given by

$$\frac{F}{A} = \eta\frac{dv}{dy} \qquad [1.2.A1.1]$$

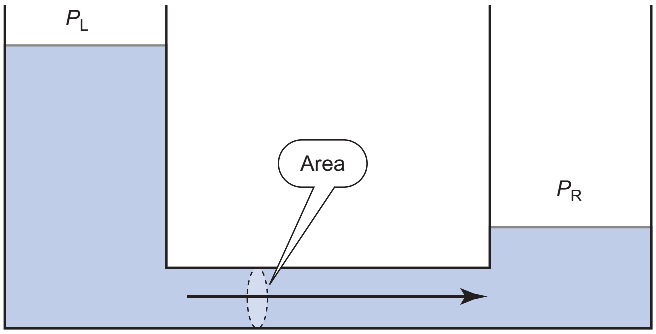

FIGURE 1.2.A1.1 Flow is driven by pressure differences.

where F is the force, A is the area, η is the coefficient of viscosity of the fluid, and dv/dy is the gradient of velocity. This equation makes good common sense: the force necessary to keep the lamina moving depends on how much area is exposed to the resistance, it depends on how sticky the fluid is, and it depends on how fast it is going.

FLOW THROUGH A PIPE

We consider here the flow through a tube or pipe of radius a and length Δx. The pressure at the left end of the pipe is $P(x)$ and the pressure at the right end is $P(x+\Delta x)$. We consider here a constant flow which therefore has a constant velocity. This means that the fluid is not subjected to any net force. Consider the forces experienced by a hollow shell of inner radius r and outer radius $r+\mathrm{d}r$, as shown in Figure 1.2.A1.2.

The fluid in contact with the walls of the pipe does not move, so that $v=0$ at $r=a$. The velocity of the fluid increases as one approaches the center of the tube. The actual velocity profile will be solved on the way to deriving an expression for the total flow through the pipe.

Because the fluid flows through the pipe at a constant velocity along the pipe (but which depends on the distance from the walls of the pipe), the sum of the forces on the hollow shell shown in Figure 1.2.A1.2 must be zero. The forces to the right include the force of the pressure on the left and the drag force of the inner layer of fluid, and the forces to the left include the pressure on the right and the drag force of the outer fluid. Thus, we have

[1.2.A1.2] $$0 = P(x)A + F_2 - P(x+\Delta x)A - F_1$$

The area of the hollow sphere on which the pressures act is $2\pi r\mathrm{d}r$. The drag forces F_1 and F_2 are given by Eqn [1.2.A1.1] as

[1.2.A1.3] $$F_1 = 2\pi(r+\mathrm{d}r)\Delta x\eta\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_{r+\mathrm{d}r}$$
$$F_2 = 2\pi r\Delta x\eta\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_{r}$$

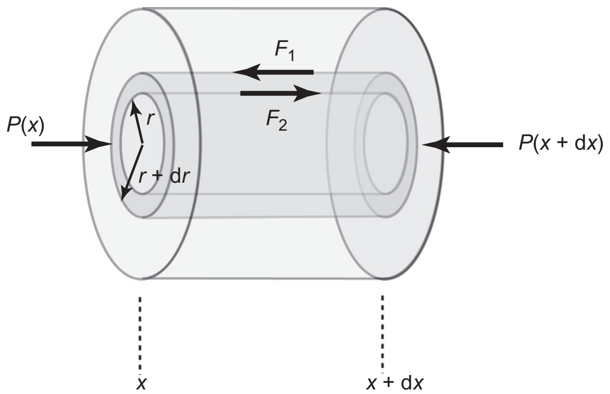

FIGURE 1.2.A1.2 Balance of forces on a hollow cylinder of inner radius r and outer radius $r+\mathrm{d}r$. F_1 is the drag force of the lamina with inner radius $r+\mathrm{d}r$ on the hollow cylinder; F_2 is the drag force on the hollow cylinder by the lamina immediately inside, with outer radius r.

Insertion of this result into Eqn [1.10.A1.2] gives

[1.2.A1.4] $$0 = (P(x) - P(x+\Delta x))2\pi r\,\mathrm{d}r + 2\pi r\Delta x\eta\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_r + 2\pi(r+\mathrm{d}r)\Delta x\eta\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_{r+\mathrm{d}r}$$

which can be re-written as

$$0 = -(\Delta P)2\pi r\,\mathrm{d}r + 2\pi r\Delta x\eta\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_r - 2\pi r\Delta x\eta\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_{r+\mathrm{d}r} - 2\pi\,\mathrm{d}r\Delta x\eta\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_{r+\mathrm{d}r}$$

[1.2.A1.5]

where $\Delta P = P(x+\Delta x) - P(x)$. We next approximate the gradient of velocity at $(r+\mathrm{d}r)$ by the first two terms of a Taylor's series:

[1.2.A1.6] $$\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_{r+\mathrm{d}r} \simeq \left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_r + \mathrm{d}r\left(\frac{\mathrm{d}^2v}{\mathrm{d}r^2}\right)_r$$

Substitution of Eqn [1.2.A1.6] into Eqn [1.2.A1.5] gives

$$0 = -\left(\frac{\Delta P}{\Delta x}\right)2\pi r\,\mathrm{d}r\Delta x + 2\pi r\Delta x\eta\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_r - 2\pi r\Delta x\eta\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_r - 2\pi\,\mathrm{d}r\,\Delta x\eta\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_r - 2\pi r\,\mathrm{d}r\Delta x\eta\left(\frac{\mathrm{d}^2v}{\mathrm{d}r^2}\right)_r - 2\pi r\,\mathrm{d}r^2\Delta x\eta\left(\frac{\mathrm{d}^2v}{\mathrm{d}r^2}\right)_r$$

[1.2.A1.7]

Canceling out the like terms, and factoring out the term $2\pi r\mathrm{d}r\ \Delta x$, we obtain

$$0 = -\left(\frac{\Delta P}{\Delta x}\right) - \frac{\eta}{r}\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right)_r - \eta\left(\frac{\mathrm{d}^2v}{\mathrm{d}r^2}\right)_r - \mathrm{d}r\frac{\eta}{r}\left(\frac{\mathrm{d}^2v}{\mathrm{d}r^2}\right)_r$$

[1.2.A1.8]

In the limit as $\mathrm{d}r\to 0$, the last term in Eqn [1.2.A1.8] vanishes. Multiplying through by -1, we are left with the differential equation

[1.2.A1.9] $$0 = \frac{\Delta P}{\Delta x} - \frac{\eta}{r}\left(\frac{\mathrm{d}v}{\mathrm{d}r}\right) + \eta\left(\frac{\mathrm{d}^2v}{\mathrm{d}r^2}\right)$$

This second-order differential equation can be solved by converting it to a first order differential equation with the substitution $y = \mathrm{d}v/\mathrm{d}r$, and then multiplying through by an integrating factor, ρ. Re-arranging Eqn [1.2.A1.9], we have

$$\rho\left(\frac{dy}{dr}\right)+\frac{\rho}{r}y+\frac{\rho}{\eta}\frac{\Delta P}{\Delta x}=0 \qquad [1.2.A1.10]$$

We choose ρ so that the first two terms are an exact differential. This is true if $\rho = r$. Thus we have

$$r\left(\frac{dy}{dr}\right)+y=\frac{d(ry)}{dr}=-\frac{r}{\eta}\left(\frac{\Delta P}{\Delta x}\right) \qquad [1.2.A1.11]$$

Integrating Eqn [1.2.A1.11], we obtain

$$\begin{aligned}\int d(ry) &= \int -\frac{r}{\eta}\left(\frac{\Delta P}{\Delta x}\right)dr\\ ry &= -\frac{1}{2\eta}\left(\frac{\Delta P}{\Delta x}\right)r^2\end{aligned} \qquad [1.2.A1.12]$$

Canceling the r factor on both sides of Eqn [1.2.A1.12] and recalling that $y = dv/dr$, we integrate Eqn [1.2.A1.12] again:

$$\begin{aligned}\frac{dv}{dr} &= -\frac{1}{2\eta}\left(\frac{\Delta P}{\Delta x}\right)r\\ \int dv &= \int -\frac{1}{2\eta}\left(\frac{\Delta P}{\Delta x}\right)r\,dr\\ v &= -\frac{1}{4\eta}\left(\frac{\Delta P}{\Delta x}\right)r^2 + C\end{aligned} \qquad [1.2.A1.13]$$

where C is a constant of integration which can be evaluated from the boundary conditions. The boundary conditions are that $v = 0$ when $r = a$. That is, the velocity of the fluid immediately adjacent to the walls of the pipe is zero. Insertion of $v = 0$ and $r = a$ into Eqn [1.2.A1.13] gives

$$\begin{aligned}C &= \frac{a^2}{4\eta}\left(\frac{\Delta P}{\Delta x}\right)\\ v &= \frac{1}{4\eta}\left(\frac{\Delta P}{\Delta x}\right)(a^2-r^2)\end{aligned} \qquad [1.2.A1.14]$$

This equation gives the velocity of fluid flow as a function of the radial distance from the center ($r = 0$) to the edge ($r = a$) of the pipe. This equation says that the velocity profile is parabolic.

What we wanted to do at the outset was to calculate the total flow through the pipe, Q_V. This is the volume of fluid which crosses the total cross-section of the pipe per unit time. The volume flux, J_V, is the volume moving through a small increment of unit area per unit time. In fact J_V is the velocity of fluid movement. To see this, consider a block of fluid moving at constant velocity, v. The block has a cross-sectional area, A. In time t the block moves a distance $v\Delta t$. The volume of fluid moving in this time is $Av\Delta t$. The volume flux, J_V, is the volume of fluid moved per unit area per unit time. This is

$$J_V = \frac{Av\Delta t}{A\Delta t} = v \qquad [1.2.A1.15]$$

Thus, the volume flux is the velocity of fluid movement. To find the total fluid flow, we integrate the flow as a function of distance from the center of the pipe. Thus

$$\begin{aligned}Q_V &= \int_0^a dQ_V = \int_0^a J_V\,dA = \int_0^a J_V 2\pi r\,dr\\ &= \int_0^a v2\pi r\,dr\\ &= \int_0^a \frac{1}{4\eta}\left(\frac{\Delta P}{\Delta x}\right)(a^2-r^2)2\pi r\,dr\\ &= \frac{2\pi}{4\eta}\left(\frac{\Delta P}{\Delta x}\right)\int_0^a (a^2-r^2)r\,dr\end{aligned} \qquad [1.2.A1.16]$$

The last integral can be evaluated by making the substitution $u = (a^2 - r^2)$, so $du = -2r dr$, to obtain

$$\begin{aligned}\int_{r=0}^{r=a}(a^2-r^2)r\,dr &= \int_{r=0}^{r=a} u\left(-\frac{du}{2}\right)\\ &= -\frac{1}{2}\int_{u=a^2}^{u=0} u\,du\\ &= \frac{a^4}{4}\end{aligned} \qquad [1.2.A1.17]$$

Inserting this result of the integration into Eqn [1.2.A1.16], we obtain

$$Q_V = \frac{\pi a^4}{8\eta}\left(\frac{\Delta P}{\Delta x}\right) \qquad [1.2.A1.18]$$

This last equation gives the total flow through the pipe, Q_V. This equation is called Poiseuille's Law, in honor of the French physician Jean Leonard Marie Poiseuille, who experimentally established the law in 1835. The relation shows that the total flow is linearly dependent on the driving force for the flow, the pressure difference between the left and right ends of the pipe. Further, it is inversely related to the length of the pipe, Δx, and to the viscosity, η. Most importantly, the flow is proportional to the fourth power of the radius of the pipe. One of the points of the derivation given above was to show the mathematical origin of this very steep dependence on the size of the pipe.

APPENDIX 1.2.A2 INTRODUCTORY STATISTICS AND LINEAR REGRESSSION

INTRODUCTION

Statistics have two functions: (1) to describe the variation in some data and (2) perform tests on sets of data to determine the cause of the variation in that data. As an example, we know that people of the same age are not the same height or weight. The set of heights for a population can be described by a statistic that describes the center (mean, median or mode) and by a statistic that describes the variation around that center. Typically the mean is used to describe the center and the standard deviation is used to describe the variation, or spread around that center. These are examples of descriptive statistics. Suppose we wanted to know the causes of variation in human height. Possible causes might include diet, genetics, sleep patterns. It is extremely difficult to determine the cause of variation in humans because these possible causes are not controlled. However, there are populations that differ in a variety of ways. We may ask the question, do two different populations have different average heights? These kinds of questions involve statistical tests.

THE MEAN

Suppose we have a set of variables $\{X_1, X_2, \ldots, X_i, \ldots, X_n\}$. The mean is ***defined*** as

[1.2.A2.1] $$\overline{X} = \mu_X = \frac{\sum_{i=1}^{n} X_i}{n}$$

THE POPULATION VARIANCE

The variance is a measure of the spread of a population. It is the average of the squared deviations from the mean. The square is taken so that all deviations from the mean contribute to the variance; otherwise, negative deviations would cancel positive deviations: the average deviation from the mean is zero. For a population the variance is given by

[1.2.A2.2] $$\sigma^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}$$

THE SAMPLE VARIANCE

The sample variance is a measure of the spread of a population when some sub-set of the population is used so that the mean is estimated rather than known. Its formula is similar to that of the population variance, except that one "degree of freedom" is used for the estimation of the mean. The sample variance is calculated as

[1.2.A2.3] $$s^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$

here s^2 is used as a symbol for the sample variance, whereas σ^2 is used as a symbol for the population variance. Note that the formula for s^2 is very similar to the formula for the variance except that $n-1$ is used as the divisor instead of n, the number of measurements used to estimate the mean and variance of the sample. The term $(n-1)$ is the **degrees of freedom** that refers to the number of independent pieces of information that goes into the estimation of the statistic. Because the mean is used in the calculations, each measurement contributes to the mean and, if you know the mean, the last value can be calculated from all of the other values. Thus the degrees of freedom when the data is used to estimate the mean is $n-1$. Expanding the square term in Eqn [1.2.A2.3], we have

[1.2.A2.4] $$s^2 = \frac{\sum_{i=1}^{n} (X_i^2 - 2\overline{X}\, X_i + \overline{X}^2)}{n-1}$$

This is rewritten as

[1.2.A2.5] $$s^2 = \frac{\sum_{i=1}^{n} X_i^2 - 2\overline{X} \sum_{i=1}^{n} X_i + n\,\overline{X}^2}{n-1}$$

Recognizing from Eqn [1.2.A2.1] that

[1.2.A2.6] $$\sum_{i=1}^{n} X_i = n\,\overline{X}$$

Inserting this into Eqn [1.2.A2.5] gives

[1.2.A2.7] $$s^2 = \frac{\sum_{i=1}^{n} X_i^2 - 2\,n\,\overline{X}^2 + n\,\overline{X}^2)}{n-1} = \frac{\sum_{i=1}^{n} X_i^2 - n\,\overline{X}^2}{n-1}$$

Multiplying both numerator and denominator by n, we reach

[1.2.A2.8] $$s^2 = \frac{n \sum_{i=1}^{n} X_i^2 - n^2\,\overline{X}^2}{n\,(n-1)}$$

Again making use of Eqn [1.2.A2.6] we arrive at the computational formula for the sample variance:

[1.2.A2.9] $$s^2 = \frac{n \sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n\,(n-1)}$$

THE SAMPLE STANDARD DEVIATION

The sample standard deviation, denoted by the statistic s, is defined to be the positive square root of the sample variance:

[1.2.A2.10] $$s = \sqrt{\frac{n \sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n\,(n-1)}}$$

THE STANDARD ERROR OF THE MEAN

The standard deviation of a sampling distribution of a statistic is called the **standard error** of that statistic. A sampling distribution of a statistic is its probability distribution. What this means is that if we estimate the mean of a population several times using different random samples, we will not get identical answers. There will be some distribution of the mean. In the same way, there will be some variation in our sample variance calculated from the different samples. The standard deviation of the distribution of means is the standard error of the mean. Note that the sample standard deviation does not change with the sample size, because each additional member of the sample contributes both to n and to the squared deviation from the mean. On the other hand, we ought to expect that our estimate of the mean improves with the number of observations in the sample. Thus, the standard error of the mean gets smaller as the number of observations in the sample increases. The formula for the standard error of the mean is

[1.2.A2.11] $$SEM = \frac{s}{\sqrt{n}}$$

PROBABILITY

Probability theory began with the analysis of games of chance. The probability of an event happening, such as drawing a particular card, or a card that belongs to a particular suit, or the appearance of a particular face on a die after a roll, is defined as the ways in which the event can happen compared to the total possible number of outcomes. For example, there are six faces on a die and the likelihood that any particular face ends upright is one out six, or $1/6 = 0.1625$. We can write this as

$$P(1) = 0.1625;\; P(2) = 0.1625;\; P(3) = 0.1625;$$
$$P(4) = 0.1625; P(5) = 0.1625;\; P(6) = 0.1625$$

where $P(i)$ refers to the probability of the face with i dots (pips) on it. If we have two dice, the probability of having a total number of dots depends on how many outcomes produce that number of dots. For example, we can find 12 dots only by having 6 on one die and 6 on the other. There are a total of 36 possible outcomes as shown in Table 1.2.A2.1 below.

Since there are 36 possible outcomes, the probability of getting a 12 is $1/36 = 0.0278$. The probability of getting a total of 7, however, is higher. We can get a 7 from the following combinations: {(4,3), (3,4), (5,2), (2,5), (6,1), and (1,6)}. The probability of getting a 7 is thus $6/36 = 0.1625$. Note the probability of getting some outcome is 1.0, and this is the sum of each of the probabilities for the individual outcomes:

$$\sum_{i=1}^{i=n} P(i) = 1.0 \qquad [1.2.A2.12]$$

where there are a total of n distinguishable outcomes. The probability of observing a particular outcome for a roll of the dice is the product of the individual outcomes for each die. Thus, the probability of a {4,3} outcome is given as

$$P(4, 3) = P(4) \times P(3) = 1/6 \times 1/6 = 1/36$$

Here we are making the distinction of which die shows 4 and which shows 3. For example, if the dice were colored this would correspond to a red die with a 4 and a white die with a 3. If they are not distinguishable, then there are two ways to get (4,3), without regard to order. In this case $P(4,3) = P(3,4)$ and the probability of getting a 7 with a 4 and a 3 is $P\{4,3\} = P(4,3) + P(3,4) = 2/36$.

The validity of these conclusions depends on **fair** dice that are **independent**. A die is fair if and only if the probabilities of all six outcomes are the same, and they are independent if and only if the outcome of one die does not influence the outcome of the other.

HYPOTHESIS TESTING

Often in scientific investigations or engineering it is necessary to obtain data to test whether or not some treatment had an effect, or whether or not the data fit some equation whose derivation is based on some theory. The derivation of the equation may have required some assumptions that limit the conditions under which the equation is valid. An excellent fit of the data to the predicted values based on the theory gives us confidence that the theory is valid and the assumptions used were met. There is no guarantee that this is so. The philosophy of science tells us that all we can do is disprove hypotheses. If the data do not agree with the theory, one of two things must be true: either there is something "wrong" with the data or there is something wrong with the theory. The data must be taken at face value but perhaps the data is not what you think it is. For example, suppose you want to investigate the relationship between pressure and flow at steady-state and you make measurements of the flow and pressure in a system. You make some measurements, but do not realize that the system is not at steady-state. Thus the data cannot be analyzed using an equation that assumes steady-state. The data are not "wrong," but you are wrong in thinking that the data represent the steady-state condition. Failure of the equation to match your data is not a problem with the data or the validity of the data, it is a problem of trying to fit data to conditions that do not pertain to how the data was actually obtained. Other possible problems exist. For example, perhaps you measured the pressure with an incorrectly calibrated meter, so the pressures that you report are not the true pressures. There is generally far more ways of making mistakes than there are ways of making the measurements correctly.

TABLE 1.2.A2.1 Possible Outcomes from a Roll of Two Dice

		Outcome of First Die					
Outcome of Second Die		1	2	3	4	5	6
	1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
	2	(1,2)	(2,2)	(3.2)	(4,2)	(5,2)	(6,2)
	3	(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
	4	(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
	5	(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
	6	(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

TABLE 1.2.A2.2 Type I and Type II Errors

Inference	Null Hypothesis	
	True	**False**
Reject H_0	Type I error	Correct inference
Accept H_0	Correct Inference	Type II error

So one kind of hypothesis is this: the derived equation is a good fit for the data. We can call this the "null hypothesis," H_0: there is no difference between the data and what is predicted from the equation. We suppose that the data is not a perfect fit. There are differences between the data and the predicted value for the data. If the differences are small, we may think that the fit is good and we are likely to accept the null hypothesis as being true. If the differences are large, we begin to think something is "wrong" with either the data or the equation, and we reject the null hypothesis as being false. It is possible to make two distinct types of errors in this testing of hypotheses. **The Type I error is the incorrect rejection of a true null hypothesis**. **The Type II error is the failure to reject a false null hypothesis**. These types of errors are clarified in Table 1.2.A2.2.

Another way of describing these is that the Type I error is a false positive: you conclude there is an effect when there is none present. A Type II error is a false negative. You fail to detect an effect when it is actually present.

The probability of making a Type I error is designated α. The probability of making a Type II error is designated as β. α is generally accepted as the level of significance of a statistical test of H_0. This is to say, we have to agree on how different the data can be from what is predicted from the null hypothesis before we can reject the null hypothesis. There is some probability that the differences that you obtained arose by random chance, and this is the probability of making a Type I error.

In other types of experiments we may want to know if some treatment or variable has some effect in a population of individuals. Suppose we do an experiment in which we measure some variable X for a set of individuals who are not treated, the control group, and we measure the same variable for a set of individuals who receive some treatment. Did the treatment affect the value of X? What we generally have is a set of measurements of variable X from the control group $G_0 = \{X_{01}, \ldots, X_{0n}\}$ and a second set of measurements from the experimental group $G_e = \{X_{E1}, \ldots, X_{En}\}$. The mean for G_0 is designated μ_0 and the mean for the experimental group is μ_E. The null hypothesis is H_0: $\mu_0 = \mu_E$. Generally the means will not be exactly equal. There is some probability that a given difference $\mu_0 - \mu_E$ will arise by chance. The t-statistic for comparison of two means is given as

$$t = \frac{\mu_0 - \mu_E}{\sqrt{\mathrm{SEM}_0^2 + \mathrm{SEM}_E^2}} \qquad [1.2.A2.13]$$

Clearly, the larger the value of t the more likely that the difference in the means does not arise by chance. If the difference arises by chance, then we would be wrong to reject the null hypothesis. The probability that the difference arises by chance, and we reject the null hypothesis, is the probability of the Type I error, and this is called the **level of significance** of the test. Typically it is accepted that $\alpha = 0.05$, which means there is a 5% chance that we will reject the null hypothesis even though it is true. Values of t for given values of α are published, and these depend on the **degrees of freedom** for the t-statistic. The degrees of freedom are the number of independent pieces of information that contribute to the estimation of the statistic, which is the t-value in this case. In a two-sample comparison such as the one described here, the degrees of freedom (symbolized as df or ν) is given as $n_0 + n_E - 2$. We subtract 2 because the means are derived from the data and so the means and the values themselves are not independent, and we have used 2 means in the calculation of the t-statistic.

We could apply these or other types of statistical test to groups of people to compare them and to make statistical statements about them, such as "men are more prone to heart attacks than women", or "black american males have a higher incidence of hypertension than white american males," or "obese white females with multiple children are most prone to gallstones." These statistical statements mean nothing when you are faced with a single incidence of the population: population averages and trends tells you nothing about any specific member of the group. You don't know, without making some measurements, how far away from average this specific person will be. However, we can make general statements about the populations that may guide us in understanding or treating a condition or setting public health policy.

THE NORMAL PROBABILITY DENSITY FUNCTION

A probability density function is one in which the probability of an outcome being in some interval is the product of the density function and the interval. Mathematically, it is given as

$$P(a \le x \le b) = \int_a^b p(x)\mathrm{d}x \qquad [1.2.A2.14]$$

where $P(a \le x \le b)$ is the probability of x being between a and b and $p(x)$ is the probability density function. There are several kinds of probability distribution functions. We will derive the normal or gaussian probability density distribution by considering a random throw of a dart aimed at the origin of a Cartesian plane. We make some basic assumptions of the probability of the dart landing in a particular area. These are:

- the distribution is symmetrical: the probability of being high by some distance is the same as being low by the same distance, and this is equal to the probability of being right by the same distance, or by being left by the same distance. In fact, the probability of being off-center by some distance, r, is independent of the angle θ from the origin.

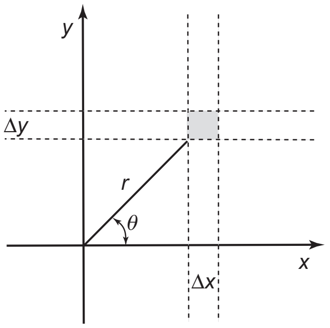

FIGURE 1.2.A2.1 Cartesian coordinate and polar coordinate systems for the analysis of the normal probability distribution function. The probability of a dart hitting the gray area, dxdy, is the probability of it landing in the vertical stripe of width Δx times the probability of it landing in the horizontal stripe with width Δy.

- errors in particular directions are independent. That is, the probability of being off-center by x in the horizontal direction does not alter the probability of being off-center by y in the vertical direction.
- large errors are less likely than small errors.
- our aim is unbiased. That is, on average, the distance from the origin, taking x and y as positive and negative, is zero.

Consider the Cartesian coordinate system shown in Fig. 1.2.A2.1. The probability of the dart landing in the vertical stripe between x and $x+\Delta x$ is given by $p(x)\Delta x$, where $p(x)$ is the probability density function. Similarly, the probability of the dart landing in the horizontal stripe between y and $y+\Delta y$ is $p(y)\Delta y$. What we want to do is determine the mathematical form of the function p.

From the independence assumption, the probability of the dart landing in the shaded region is the product of the probabilties of landing in the horizontal or vertical stripes: $p(x)\Delta x p(y)\Delta y$. Because of symmetry, the probability of the dart landing in the area the size of $\Delta x\Delta y$ located r distance from the origin does not depend on θ and we can write

[1.2.A2.15] $$g(r)\Delta x\Delta y = p(x)\Delta x\, p(y)\Delta y$$

which gives

[1.2.A2.16] $$g(r) = p(x)\, p(y)$$

Differentiating with respect to θ, we obtain

[1.2.A2.17] $$\frac{dg(r)}{d\theta} = 0 = p(x)\frac{dp(y)}{d\theta} + p(y)\frac{dp(x)}{d\theta}$$

$dg(r)/d\theta = 0$ because $g(r)$ is independent of θ. Inserting $x = r\cos\theta$ and $y = r\sin\theta$, we can re-write Eqn [1.2.A2.17] as

[1.2.A2.18] $$0 = p(x)\frac{dp(y)}{dy}\frac{dr\sin\theta}{d\theta} + p(y)\frac{dp(x)}{dx}\frac{dr\cos\theta}{d\theta}$$
$$0 = p(x)\frac{dp(y)}{dy}\, r\cos\theta - p(y)\frac{dpx}{dx} r\sin\theta$$

[1.2.A2.19] $$\frac{\dfrac{dp(x)}{dx}}{p(x)x} = \frac{\dfrac{dp(y)}{dy}}{p(y)y}$$

This differential equation is true for all x and y, and x and y are independent. This can happen only if the ratio defined on each side of the equation is constant. We let

[1.2.A2.20] $$\frac{\dfrac{dp(x)}{dx}}{p(x)x} = C$$
$$\frac{dp(x)}{p(x)} = C\,x\,dx$$
$$\int\frac{dp(x)}{px} = \int C\,x\,dx$$
$$\ln p(x) = C\frac{x^2}{2} + c$$

Taking the exponent of e of both sides of the last equation gives

[1.2.A2.21] $$p(x) = Ae^{\frac{k}{2}x^2}$$

Because large x is less likely than small x, we know that C must be negative and we can re-write this equation as

[1.2.A2.22] $$p(x) = Ae^{-\frac{k}{2}x^2}$$

We can evaluate A by the requirement that the total probability for all outcomes is 1.0. This is expressed mathematically as

[1.2.A2.23] $$\int_{-\infty}^{\infty} Ae^{-\frac{k}{2}x^2} dx = 1.0$$

Since A is a constant, it can be removed from the integrand. Due to the symmetry of the problem, this integral is twice the integral from zero to infinity. We write this as

[1.2.A2.24] $$\int_0^{\infty} e^{-\frac{k}{2}x^2}\; dx = \frac{1}{2A}$$

We can combine this with the distribution about y, as these are symmetrical:

[1.2.A2.25] $$\int_0^{\infty} e^{-\frac{k}{2}x^2} dx \cdot \int_0^{\infty} e^{-\frac{k}{2}y^2} dy = \frac{1}{4A^2}$$

Since x and y are independent, we can re-write this product of integrals as the integral of their products:

[1.2.A2.26] $$\int_0^{\infty}\int_0^{\infty} e^{-\frac{k}{2}(x^2+y^2)}\; dx\,dy = \frac{1}{4A^2}$$

This can be converted to polar coordinates, recognizing that $(x^2+y^2) = r^2$ and $dxdy = rdrd\theta$. We obtain

[1.2.A2.27] $$\int_0^{\frac{\pi}{2}}\int_0^{\infty} e^{-\frac{k}{2}r^2} r\,dr\,d\theta = \frac{1}{4A^2}$$

We can evaluate the integrals by making the substitution $u = -kr^2/2$. It follows that $\mathrm{d}u = -kr\mathrm{d}r$, and thus $-\mathrm{d}u/k = r\mathrm{d}r$. Making these substitutions in the interior integral, we derive

[1.2.A2.28] $$\int_0^{\frac{\pi}{2}} \frac{-1}{k}\left[\int_0^{-\infty} \mathrm{e}^{u}\, \mathrm{d}u\right] \mathrm{d}\theta = \frac{1}{4A^2}$$

Evaluation of the interior integral is -1. Eqn [1.2.A2.28] thus becomes

[1.2.A2.29] $$\int_0^{\frac{\pi}{2}} \frac{1}{k} \mathrm{d}\theta = \frac{\pi}{2k} = \frac{1}{4A^2}$$

The value of A is thus given as

[1.2.A2.30] $$A = \sqrt{\frac{k}{2\pi}}$$

The probability distribution in Eqn [1.2.A2.22] becomes

[1.2.A2.31] $$p(x) = \sqrt{\frac{k}{2\pi}}\,\mathrm{e}^{-\frac{k}{2}x^2}$$

We can evaluate k from the variance of the probability distribution. The average value for x is given as

[1.2.A2.32] $$\mu = \int_{-\infty}^{\infty} x\, p(x)\, \mathrm{d}x$$

The variance, similar to our earlier definition in Eqn [1.2.A2.2] is given as

[1.2.A2.33] $$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2\, p(x)\, \mathrm{d}x$$

Because of the symmetry of the coordinates, and the fact that $p(x) = p(-x)$, we know that the mean, μ, is zero. With Eqn [1.2.A2.31], and $\mu = 0$, the variance is given as

[1.2.A2.34] $$\sigma^2 = 2\sqrt{\frac{k}{2\pi}} \int_0^{\infty} x^2\, \mathrm{e}^{-\frac{k}{2}x^2}\, \mathrm{d}x$$

We use symmetry as we did before to integrate from 0 to ∞:

[1.2.A2.35] $$\sigma^2 = 2\sqrt{\frac{k}{2\pi}} \int_0^{\infty} x^2\, \mathrm{e}^{-\frac{k}{2}x^2}\, \mathrm{d}x$$

We can integrate this by parts by identifying

[1.2.A2.36] $$u = x \quad \mathrm{d}v = x\,\mathrm{e}^{-\frac{k}{2}x^2}\,\mathrm{d}x \quad v = -\frac{1}{k}\mathrm{e}^{-\frac{k}{2}x^2}$$

Substituting these into Eqn [1.2.A2.35] we obtain

$$\sigma^2 = 2\sqrt{\frac{k}{2\pi}} \int_0^{\infty} x^2\, \mathrm{e}^{-\frac{k}{2}x^2}\, \mathrm{d}x$$

$$\sigma^2 = 2\sqrt{\frac{k}{2\pi}} \int_0^{\infty} u\mathrm{d}v$$

$$\sigma^2 = 2\sqrt{\frac{k}{2\pi}} \left[uv\Big|_0^{\infty} - \int_0^{\infty} v\mathrm{d}u \right]$$

$$\sigma^2 = 2\sqrt{\frac{k}{2\pi}} \left[-\frac{x}{k}\mathrm{e}^{-\frac{k}{2}x^2}\Big|_0^{\infty} + \int_0^{\infty} \frac{1}{k}\mathrm{e}^{-\frac{k}{2}x^2}\, \mathrm{d}x \right]$$

[1.2.A2.37]

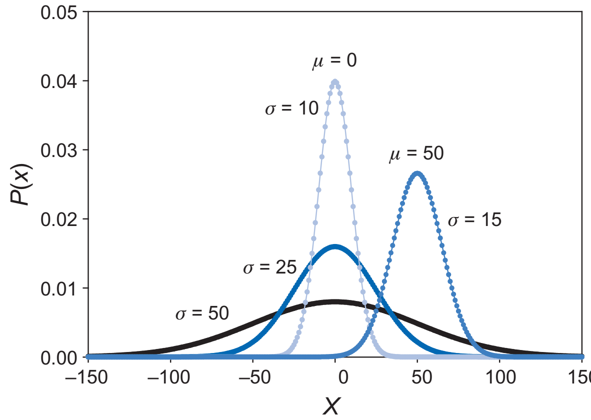

FIGURE 1.2.A2.2 Gaussian probability distribution functions for varying values of σ and μ. Different values of μ move the distribution to the right or to the left; different values of σ influence the shape of the curve, with smaller values resulting in sharper curves and larger values creating more spread out curves.

The first term in brackets is evaluated from $x = 0$ to $x = M$ in the limit as $M \to \infty$, and it is zero. The second term in the brackets has already been done in this derivation (Eqn [1.2.A2.24] and Eqn [1.2.A2.30]). This last equation in Eqn [1.2.A2.37] becomes

[1.2.A2.38] $$\sigma^2 = 2\sqrt{\frac{k}{2\pi}}\left[0 + \frac{1}{k}\frac{\sqrt{2\pi}}{2\sqrt{k}}\right] = \frac{1}{k}$$

This last equation gives $k = 1/\sigma^2$. Inserting this into the probability density function (Eqn. [1.2.A2.31]) gives

[1.2.A2.39] $$p(x) = \frac{1}{\sigma^2\sqrt{2\pi}}\mathrm{e}^{-2\sigma^2}$$

This is the normal probability distribution centered at $\mu = 0$. The general equation with mean μ is achieved by a horizontal shift in x:

[1.2.A2.40] $$p(x) = \frac{1}{\sigma^2\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\,\sigma^2}}$$

Calculated Gaussian or normal probability distribution functions are shown in Figure 1.2.A2.2. Note that the value of σ determines the spread of the distribution. Smaller values of σ result in a sharper distribution.

LINEAR REGRESSION

For linear regression, the data we're interested in is generally numerical and comes in sets of ordered pairs: $\{(x_1,y_1), \ldots, (x_n,y_n)\}$. What we desire to know is what is the best linear fit to the ordered pairs? There are many different ways of doing this. What we are going to go over is the **least squares linear regression**. The first thing we are going to do is to make a few assumptions. These are:

1. there is no error in the values of X_i. All of the error is in the values of Y_i
2. the values of Y_i are distributed normally. That is, they obey the normal probability distribution. This is the same as the Gaussian probability distribution and is commonly referred to as the "bell curve". This curve has been described in Eqn [1.2.A2.40]

We introduce a model equation

[1.2.A2.41] $$\hat{Y} = m\,X = b$$

where $\hat{Y}$ is the predicted value of Y by this model equation. Since we have assumed that the values of X are determined perfectly, with no error, all of the error is in the set of Y_i and how it differs from the predicted values, $\hat{Y}_i$. What we want to do is minimize the squared error between the set of Y_i and $\hat{Y}_i$. The total squared error is

[1.2.A2.42] $$\text{Error} = \sum_i (Y_i - \hat{Y}_i)^2$$

Substituting in from Eqn [1.2.A2.41] we have

[1.2.A2.43] $$\text{Error} = \sum_i (Y_i - mX_i - b)^2$$

The error estimated as the square of the deviations from the observed values, the set of Y_i, from the predicted values, $mX_i + b$, varies with m and b used in the calculations. We require this sum of square errors to be minimized in order to have the best fit line. We achieve this by looking for minima in the square errors as we vary m and b. That is, we look for the m and b that has the least squared error. This occurs when the partial derivatives of the error with respect to m and b are minima:

[1.2.A2.44] $$\frac{\partial \sum_i (Y_i - mX_i - b)^2}{\partial m} = 0$$
$$\frac{\partial \sum_i (Y_i - mX_i - b)^2}{\partial b} = 0$$

These two form a pair of simultaneous equations in m and b; all of the pairs of (x_i, y_i) are known. Expanding the square term, we get

$$\frac{\partial \sum_i (Y_i^2 - 2mY_iX_i - 2Y_ib + m^2X_i^2 + 2mbX_i + b^2)}{\partial m} = 0$$
$$\frac{\partial \sum_i (Y_i^2 - 2mY_iX_i - 2Y_ib + m^2X_i^2 + 2mbX_i + b^2)}{\partial b} = 0$$

[1.2.A2.45]

Performing the partial differentiation gives

[1.2.A2.46] $$2m\sum_i X_i^2 + 2b\sum_i X_i - 2\sum_i Y_iX_i = 0$$
$$2m\sum_i X_i + 2nb - 2\sum_i Y_i = 0$$

canceling out the common factors of 2, this is recognized as a system of two simultaneous equations in two unknowns:

[1.2.A2.47] $$\sum_i X_i^2 m + \sum_i X_i\, b = \sum_i Y_iX_i$$
$$\sum_i X_i m + n\,b = \sum_i Y_i$$

This system of simultaneous equations in two unknowns (m and b) can be solved by application of Cramer's Rule:

[1.2.A2.48] $$\frac{\begin{vmatrix} \sum_i X_iY_i & \sum_i X_i \\ \sum_i Y_i & n \end{vmatrix}}{\begin{vmatrix} \sum_i X_i^2 & \sum_i X_i \\ \sum_i X_i & n \end{vmatrix}} = m$$

and the expression for b is

[1.2.A2.49] $$\frac{\begin{vmatrix} \sum_i X_i^2 & \sum_i X_iY_i \\ \sum_i X_i & \sum_i Y_i \end{vmatrix}}{\begin{vmatrix} \sum_i X_i^2 & \sum_i X_i \\ \sum_i X_i & n \end{vmatrix}} = b$$

Evaluating the determinants in the numerator and denominator for m in Eqn [1.2.A2.48], we get the computational formula for the slope of the least-squares best fit line:

[1.2.A2.50] $$m = \frac{n\sum_i X_iY_i - \sum_i X_i \sum_i Y_i}{n\sum_i X_i^2 - \sum_i X_i \sum_i X_i}$$

Evaluating the determinants from Eqn [1.2.A2.49] we get

[1.2.A2.51] $$b = \frac{\sum_i X_i^2 \sum_i Y_i - \sum_i X_i \sum_i X_iY_i}{n\sum_i X_i^2 - \sum_i X_i \sum_i X_i}$$

By algebraic manipulation, it can be shown that this last equation for the intercept of the least-squares line is given also as

[1.2.A2.52] $$b = \overline{Y} - m\overline{X}$$