Week 7 notes: other continuous distributions : **Beta, Weibull ,**
Joint distributions, Covariance, Conditional distributions, normal scores plot

We continued doing in class some of the examples on exponential random variables written up in last week's notes. We also noted that in the EXAMPLE 7 d) from last week's notes that the probability that our exponential waiting time with parameter $\lambda = .25\,, \beta = 4 = 1/\lambda$ exceeds 8 minutes is the same as the probability that zero events have occurred in the underlying Poisson process Y with parameter $\lambda t = (.25)t$ with t = 8 giving rise to our exponential waiting time $P(X > 8) = P(Y(8) = 0) = e^{-(.25)8}$. This is in fact how one shows that the waiting times between Poisson events are exponentially distributed.

Beta distribution . This is the probability distribution with density

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

Recall that the Gamma function is a generalization of the factorial function from integer to non-integer values with $\Gamma(\alpha) = (\alpha-1)!$ when α is an integer. Thus the coefficient out in front of the beta distribution resembles a binomial coefficient (but not exactly) and the whole density resembles the density of a binomial random variable except now the probability $x = p$ of success is unknown and has a beta probability distribution placed upon it.

Note that the **uniform distribution** results as the special case where $\alpha = 1\,, \beta = 1$ We have not talked about the generalization of **Bayes formula for continuous random variables** but suffice to say it can be done. In the Bayesian context the beta distribution can be regarded as the prior distribution for a binomial random variable when the success probability p is unknown. Then after we obtain some data we update our knowledge via Bayes' formula getting a posterior distribution. The beta distribution has the nice property that this posterior is also a beta distribution but the parameters change depending on what data is observed.

In the same way that a multinomial distribution is a generalization of the binomial distribution, the beta distribution has a higher dimensional generalization known as the **Dirichlet distribution** that goes along with a general multinomial distribution. It too has the nice property that if the prior is Dirichlet so is the posterior (but with updated parameters) .

Relation to the Gamma distribution : The beta distribution is also related to the Gamma distribution and is the distribution of a quotient X/(X+Y) where X and Y are independent Gamma random variables with parameters $(\alpha . \theta)$ and $(\beta\,, \theta)$ respectively

Weibull distribution $f(x) = \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}}$ for $x > 0, \alpha > 0\,, \beta > 0$ (= 0 else)

We remarked that if a random variable ***X*** has the Weibull distribution with parameters $\alpha\,, \beta$ then the random variable $Y = X^{\beta}$ will be exponentially distributed with parameter $\lambda = \alpha$. The above density can be recovered from that of Y namely $f_Y(y) = \alpha e^{\alpha y}$ by the chain rule via

$$f_X(x) = \frac{d}{dx} P(X \le x) = \frac{d}{dx} P(Y = X^{\beta} \le y = x^{\beta}) = \frac{d}{dy} F_Y(y) \frac{dy}{dx} = f_Y(y) \beta x^{\beta-1} = \alpha e^{\alpha y} \beta x^{\beta-1}$$

upon replacing y by x^{β} .

Joint probability distributions – discrete case . We have already seen a joint probability mass function (pmf) in our discussion of the multinomial distribution for which we derived the joint pmf.

EXAMPLE 1 Now let us consider another example in two dimensions. Suppose we randomly sample two components from a large batch in which 1/4 of them are defective. If either we sample with replacement or the population we sample from is sufficiently large so that the probability p = 1/4 of selecting a defective does not change with each selection, and if each selection is independent of other selections then consider the random variables

Y = the number of defectives in our sample of size 2 (so Y has possible values 0, 1, or 2)

X = 1 if the first item sampled is defective, else X = 0 (so X has possible values 0 or 1)

Letting N = non-defective, D= defective there are 4 possible outcomes :

(N, N) or X=0, Y=0 with probability $(3/4)^2=9/16$

(N, D) or X=0, Y=1 with probability $(3/4)(1/4)=3/16$

(D, N) or X=1, Y=1 with probability $(1/4)(3/4)=3/16$

(D, D) or X=1, Y=2 with probability $(1/4)^2=1/16$

a) We summarize our findings in a table for the **joint pmf** for the random variables X and Y

$p_{X,Y}(x,y)=p(x,y)=P(X=x,Y=y)$

		Y = 0	Y = 1	Y = 2	total = $p_X(x)$
X	0	9/16	3/16	0	12/16 = 3/4
	1	0	3/16	1/16	4/16 = 1/4
total = $p_Y(y)$		9/16	6/16	1/16	1

The row sums where for fixed x we sum over all y gives the **marginal pmf for *X* :**

$$p_X(x)=\sum_y p(x,y)$$

This is just a statement of the law of total probability since for any x , the disjoint events $\{Y=y\}$ partition the event $\{X=x\}$ into a disjoint union so that probabilities add :

$$p_X(x)=P(X=x)=P(\cup_y\{X=x\}\cap\{Y=y\})=\sum_y P(X=x,Y=y)=\sum_y p(x,y)$$

Similarly the column sums where we fix y and sum over x gives the **marginal pmf for *Y*** :

$$p_Y(y)=\sum_x p(x,y)$$

b) We can now compute $\mu_X=E[X]=\sum_x x\,p_X(x)=0\cdot 3/4+1\cdot 1/4=1/4$

$E[X^2]=\sum_{x^2} x\,p_X(x)=0\cdot 3/4+1\cdot 1/4=1/4$ so that $V[X]=E[X^2]-\mu_X^2=1/4-(1/4)^2=3/16$

and similarly $\mu_Y=E[Y]=\sum_y y\,p_Y(y)=0\cdot 9/16+1\cdot 6/16+2\cdot 1/16=1/2$

$E[Y^2]=\sum_{y^2} y\,p_Y(y)=0^2\cdot 9/16+1^2\cdot 6/16+2^2\cdot 1/16=5/8$ giving for the variance of Y

$V[Y]=E[Y^2]-\mu_Y^2=5/8-(1/2)^2=3/8$

We can also compute the probability mass function of the **conditional distribution of *X* given *Y*** by

$$p_{X|Y}(x|y)=p_1(x|y)=P(X=x|Y=y)=\frac{p(x,y)}{p_Y(y)}$$

The subscript 1 is to indicate it is a distribution in the first variable X (we could have also asked for the conditional distribution $p_2(y|x)$ of Y given X)

c) For the above example we have $p_1(0|0)=\dfrac{p(0,0)}{p_Y(0)}=\dfrac{9/16}{9/16}=1$, which implies $p_1(1|0)=0$

$$p_1(0|1)=\frac{p(0,1)}{p_Y(1)}=\frac{3/16}{6/16}=\frac{1}{2}\text{, which implies } p_1(1|1)=\frac{1}{2} \text{ and}$$

$$p_1(0|2)=\frac{p(0,2)}{p_Y(0)}=\frac{0}{1/16}=0$$, which implies $p_1(1|2)=1$

Note that for example $p_1(0|2)=P(X=0|Y=2)=0\neq P(X=0)=p_1(0)=3/4$. That is the random variables *X* and *Y* are **not independent** since independence would mean that conditioning doesn't change any of the probabilities for *X*. Equivalently we conclude that *X* and *Y* are not independent since the joint probability does not factor in all cases :

$$p_{X,Y}(0,2)=P(X=0,Y=2)=0\neq P(X=0)P(Y=2)=P_X(0)P_Y(2)=(3/4)(1/16)=3/64$$

Two discrete random variables *X* and *Y* are **independent** means that their joint pmf factors as

$$p_{X,Y}(x,y)=p_X(x)\,p_Y(y) \quad \text{for all values of } x \text{ and } y$$

Note for any joint probability mass function if *A* denotes a set of points (*x*,*y*) in the plane, then the event

$(X,Y)\in A$ is the disjoint union of the events

$(X,Y)=(x,y)$ or equivalently $\{X=x\}\cap\{Y=y\}$ for all possible $(x,y)\in A$

Consequently

$$P((X,Y)\in A)=\sum_{(x,y)\in A} p(x,y) \ .$$

If we define the **joint cumulative distribution function**

$$F_{X,Y}(x,y)=P(X\leq x,Y\leq y)=\sum_{s\leq x,\ t\leq y} p(s,t)$$

independence can also be expressed in terms of the cdf as the condition that

$$F_{X,Y}(x,y)=F_X(x)F_Y(y) \text{ for all } x,y$$

since if *X* and *Y* are independent the joint pmf factors so that we have

$$F_{X,Y}(x,y)=\sum_{s\leq x}\sum_{t\leq y} p_X(s)\,p_Y(t)=\sum_{s\leq x} p_X(s)\sum_{t\leq y} p_Y(t)=P(X\leq x)P(Y\leq y)$$

The expected value of any function $h(X,Y)$ of *X* and *Y* **:** We now ask what is meant by the expected value $E[XY]$ corresponding to the choice $h(x,y)=xy$ or more generally any (measurable) function *h* of *x* and *y* . As in the case of a single variable where we expressed an expectation of a function $h(X)$ in terms of the pmf of the original variable *X* we are lead to the definition in terms of the joint pmf for the original random variables *X* and *Y :*

$$E[h(X,Y)]=\sum_{x,y} h(x,y)\,p_{X,Y}(x,y)$$

d) Thus for the above example

$$E[XY]=\sum_{x,y} xy\,p(x,y)=0+1\cdot 1\cdot 3/16+1\cdot 2\cdot 1/16=5/16$$

from which we can compute the *covariance*

$$cov[X,Y]=E[XY]-\mu_X\,\mu_Y=5/16-(1/4)(1/2)=3/16 \quad \text{of } X \text{ and } Y.$$

The (population) **covariance** of *X* and *Y* is a measure of their joint variation from their means and for both discrete and continuous random variables it is defined as

$$cov[X,Y]=cov[Y,X]=E[(X-\mu_X)(Y-\mu_Y)]=E[XY]-\mu_X\,\mu_Y$$

where the last equality follows by algebra. This reduces to the usual formula for the variance of *X* if we replace *Y* by *X* and to the variance of *Y* if we replace *X* by *Y* :

$$V[X]=cov[X,X] \quad \text{and} \quad V[Y]=cov[Y,Y] \ .$$

Often the covariance is measured relative to a natural length scale determined by the standard deviations of *X* and *Y* . This gives rise to the **correlation** between *X* and *Y* defined as

$$\rho(X,Y)=\frac{cov[X,Y]}{\sigma_X\sigma_y}=E[\frac{(X-\mu_X)}{\sigma_X}\frac{(Y-\mu_Y)}{\sigma_Y}]\ .$$

One can show the correlation lies between -1 and 1 : $-1\le\rho(X,y)\le 1$.

Given n pairs of data values (X_i,Y_i) one also can speak of a **sample covariance** and **sample correlation**

$$s_{xy}=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{x})(Y_i-\bar{y}) \text{ and } \rho_{xy}=\frac{1}{n-1}\sum_{i=1}^{n}\frac{(X_i-\bar{x})}{s_X}\frac{(Y_i-\bar{y})}{s_Y}$$

which reduces to the sample variance s_X^2 if we replace Y_i by X_i and to the sample variance s_Y^2 if we replace X_i by Y_i .

(One needs to be careful not to confuse this notation for the sample covariance with a similar notation commonly used for the sum $S_{xy}=\sum_{i=1}^{n}(X_i-\bar{x})(Y_i-\bar{y})$ which differs by the factor of 1/(n-1) .)

For a random vector in k dimensions $\vec{X}=(X_1,X_2,\ldots,X_k)^T$ (which we regard as a column vector) define the expectation $E[\vec{X}]=(E[X_1],E[X_2],\ldots,E[X_k])^T$ (a column vector) and similarly for the expectation of a random matrix, the matrix of expected values of its components, then the $k\times k$ **covariance matrix** of the random vector $\vec{X}$ is the matrix $cov[\vec{X}]=E[(\vec{X}-E[\vec{X}])(\vec{X}-E[\vec{X}])^T]$ $=E[\vec{X}\vec{X}^T]-E[\vec{X}]E[\vec{X}]^T$ which has (i, j) entry $cov[X_i,X_j]$. Note it is **symmetric** : ((i, j) entry = $cov[X_i,X_j]=cov[X_j,X_i]$ = (j, i) entry) . Note that along the diagonal of the covariance matrix are the variances of the different components of the random vector.

e) In the above example with k=2, $(X_1,X_2)=(X,Y)$ which is to say $X_1=X$ and $X_2=Y$ and the covariance matrix is a 2 by 2 matrix with entries the variance $V[X]=3/16$ and the variance $V[Y]=3/8$ along the diagonal and the two off diagonal elements have the same value

$$cov[X,Y]=E[XY]-\mu_X\mu_Y=3/16\ .$$

EXAMPLE 2 (like problem 5.71 of text) Two scanners are needed for an experiment. Of the 8 available,
3 have electronic defects , 2 have memory defects and 3 are in good working order.

Let X_1 = the number with electronic defects (possible values = 0, 1 or 2)
X_2 = number with defect in memory (possible values = 0, 1 or 2)

a) Find the joint distribution :

$$p(j,k)=\binom{3}{j}\binom{2}{k}\binom{3}{2-(j+k)}\ /\ \binom{8}{2} \text{ for } 0\le j+k\le 2$$

b) Find the probability that there are 0 or 2 total defects among the 2 scanners selected

$$P(X_1+X_2=0 \text{ or } X_1+X_2=2)=p(0,0)+p(1,1)+p(0,2)+p(2,0)=(3+6+1+3)/28=13/28$$

c) Find the marginal distribution of X_1 :

$$p_1(j)=\sum_{k=0}^{2-j}p(j,k) \text{ so } p_1(0)=p(0,0)+p(0,1)+p(0,2)=10/28\ ;$$
$$p_1(1)=p(1,0)+p(1,1)=15/28 \text{ and } p_1(2)=p(2,0)=3/28$$

d) Find the conditional distribution of X_1 given X_2 :

$$P(X_1=j\,|\,X_2=0)=p(j,0)/p_2(0)=p(j,0)/(p(0,0)+p(1,0)+p(2,0))$$
$$p(0|0)=p(0,0)/p_2(0)=p(0,0)/(p(0,0)+p(1,0)+p(2,0))=1/5\ ;\ \ p(1|0)=3/5\ ;\ \ p(2|0)=1/5$$

EXAMPLE 3 (like 5.72) Two random variables are independent, each with binomial distribution with success probability p = .4 and n=3 trials

a) Find the joint distribution : $p(j,k)=p_1(j)\,p_2(k)=\binom{3}{j}\binom{3}{k}(.4)^{j+k}(.6)^{6-(j+k)}$ for $0\le j\le 3,\ 0\le k\le 3$

b) Find the probability that the second random variable is greater than the first :

$$p(X_2>X_1)=p(0,1)+p(0,2)+p(0,3)+p(1,2)+p(1,3)+p(2,3)$$

$$=3\cdot\left[\ (.4)(.6)^5+(.4)^2(.6)^4+(.4)^4(.6)^2+(.4)^5(.6)\ \right]+10(.4)^3(.6)^3$$

Expectation and Variance of a linear combination : (discrete random variables case)

Let us also examine the expected value of the sum $E[aX+bY]$ for any constants a,b

$$E[aX+bY]=\sum_x\sum_y(ax+by)\,p(x,y)=a\sum_x x\sum_y p(x,y)+b\sum_y y\sum_x p(x,y)$$

$$=a\sum_x x\,p_X(x)+b\sum_y y\,p_Y(y)=a\,E[X]+b\,E[Y]$$

Note : for the expected value we do not need to assume independence of *X* and *Y*.

For the variance calculation however we do assume independence of *X* and *Y* : then

$$V[aX+bY]=\sum_x\sum_y(a(x-\mu_X)+b(y-\mu_Y))^2\,p_1(x)\,p_2(y)\ =$$

$$a^2\sum_x(x-\mu_X)^2\,p_1(x)+b^2\sum_y(y-\mu_Y)^2\,p_2(y)=a^2V[X]+b^2V[Y]$$

using the fact that the marginals sum to 1 and that the cross term

$$2ab\sum_x(x-\mu_X)\,p_1(x)\sum_y(y-\mu_Y)\,p_2(y)=2\text{abE}[X-\mu_X]E[Y-\mu_Y]$$

vanishes since

$$E[X-\mu_X]=0=E[Y-\mu_Y]$$

Remark : The situation for 3 or more dimensions is similar to the two dimensional case. One finds the marginal distribution for one variable by summing the joint pmf over all other variables.

Joint density for continuous random variables : A joint density $f(x_1,x_2,\ldots,x_k)$ in *k* continuous random variables $X_1,X_2,\ldots,X_k$ must satisfy

1) $f(x_1,x_2,\ldots,x_k)\ge 0$
2) $\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,x_2,\ldots,x_k)\,dx_1\,dx_2\ldots dx_k\ =\ 1$ (total probability is 1)
3) $P(a_1\le X_1\le b_1,a_2\le X_2\le b_2,\ldots a_k\le X_k\le b_k)=\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_k}^{b_k}f(x_1,x_2,\ldots,x_k)\,dx_k\ldots dx_2\,dx_1$

This last condition is the defining property of the joint density. It says that the probability that the random vector $\vec{X}=(X_1,X_2,\ldots,X_k)$ lies in the rectangle $A=\{a_1\le x_1\le b_1,a_2\le x_2\le b_2,\ldots,a_k\le x_k\le b_k\}$ is just the integral of the joint density over this rectangular region. The other two properties are then really just consequences of the axioms of probability. By the countable additivity of probabilities and by the additivity of the integral, for any *k*-dimensional set *A* which is a countable disjoint union of rectangles and this includes almost any set *A* of interest, not just for rectangles, the probability that the random vector $\vec{X}=(X_1,X_2,\ldots,X_k)$ lies in the set *A* is just the integral of the joint density over the region *A* :

$$P((X_1,X_2,\ldots,X_k)\in A)=\int_A\cdots\int f(x_1,x_2,\ldots,x_k)\,dx_1\,dx_2\ldots dx_k\ .$$

The **marginal density** for any of the variables is obtained by integrating out the other variables, so that for example the marginal for X_1 is $f_1(x_1)=f_{X_1}(x_1)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f(x_1,x_2,\ldots,x_k)\,dx_2\,dx_3\ldots dx_k$

The **joint cumulative distribution function (cdf)** is

$$F(x_1, x_2, \ldots, x_k) = P(X_1 \le x_1, X_2 \le x_2, \ldots, X_k \le x_k) = \int_{-\infty}^{x_1} \cdot \cdot \int_{-\infty}^{x_k} f(s_1, s_2, \ldots, s_k)\, ds_1\, ds_2 \ldots ds_k$$

Continuous random variables are **independent** if either the joint density factors as the product of the marginal densities :

$$f(x_1, x_2, \ldots, x_k) = f_1(x_1)\, f_2(x_2) \cdot \cdot \cdot f_k(x_k)$$

or equivalently the joint cumulative distribution factors as the product of the cumulative distributions corresponding to each marginal density :

$$F(x_1, x_2, \ldots, x_k) = F_1(x_1) \cdot \cdot \cdot F_k(x_k) \quad .$$

The **expected value** of any function $h(X_1, X_2, \ldots, X_k)$ of the random variables $X_1, X_2, \ldots, X_k$ is given by

$$E[h(X_1, X_2, \ldots, X_k)] = \int_{-\infty}^{\infty} \cdot \cdot \int_{-\infty}^{\infty} h(x_1, x_2, \ldots, x_k)\, f(x_1, x_2, \ldots, x_k)\, dx_1\, dx_2 \ldots dx_k$$

Conditional densities are a little trickier to define since if we use the definition of conditional probability the event $X_2 = x_2$ that we condition on, has $P(X_2 = x_2) = 0$ i.e. 0 probability. But if we interpret this in a limiting sense we can make sense out of it this way : for very small interval widths Δx_1 and Δx_2 the conditional probability that the first variable lies in a small interval given that the second does is

$$P(X_1 \in [x_1, x_1 + \Delta x_1] \;\mid\; X_2 \in [x_2, x_2 + \Delta x_2]) = \frac{P(X_1 \in [x_1, x_1 + \Delta x_1], X_2 \in [x_2, x_2 + \Delta x_2])}{P(X_2 \in [x_2, x_2 + \Delta x_2])}$$

is approximately $\approx \dfrac{f(x_1, x_2)\, \Delta x_1\, \Delta x_2}{f_2(x_2)\, \Delta x_2} = f(x_1 | x_2)\, \Delta x_1$.

This approximation becomes exact in the limit as $\Delta x_1 \to 0$ and $\Delta x_2 \to 0$. This motivates the

Definition : The **conditional probability density** of X_1 given X_2 is

$$f_{X_1|X_2}(x_1 | x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}$$

provided the denominator is not 0 (the marginal density for the second variable is non-zero)
Note that the marginal density at a point does not vanish so that the original difficulty has gone away even though we still have $P(X_2 = x_2) = 0$.

EXAMPLE 4 (like 5.73, 5.74, 5.75) (You might want to look at my comments on how to do multiple integrals given in EXAMPLE 5 below. One does the inner integral first, treating occurrences of the outer variable as constant within the integrand, with the limits of integration for the inner integral possibly depending on the outer variable as in part c) of this example below)

If two random variables have the joint density

$$f(x, y) = kxy^2 \text{ for } 0 < x < 1,\; 0 < y < 2 \quad (= 0 \text{ elsewhere })$$

a) Find the constant *k* that makes this a density : the total probability equals 1 so

$$k \int_0^1 \int_0^2 x\, y^2\, dy\, dx = k \int_0^1 (8/3)\, x\, dx = 4\mathrm{k}/3 = 1 \text{ gives } k = 3/4 \quad .$$

b) Find the probability that both random variables will take values less than 1 :

$$P(X < 1, Y < 1) = \int_0^1 \int_0^1 (3/4)\, x\, y^2\, dy\, dx = \int_0^1 (1/4)\, x\, dx = 1/8$$

c) Find the probability that the sum of the two random variables will be less than 1 :

$$P(X + Y < 1) = \int_0^1 \int_0^{1-x} (3/4)\, x\, y^2\, dy\, dx = \int_0^1 (1/4)\, x (1-x)^3 dx = (1/4) \int_0^1 x - 3\mathrm{x}^2 + 3x^3 - x^4\, dx = 1/80$$

d) Find the marginal densities

$$f_1(x)=\int_0^2 (3/4)x\,y^2\,dy=2\mathrm{x} \text{ for } 0<x<1 \quad (=0 \text{ else })$$

$$f_2(y)=\int_0^1 (3/4)x\,y^2\,dx=(3/8)\,y^2 \text{ for } 0<y<2 \quad (=0 \text{ else })$$

Are the random variables independent ? The joint density factors as the product of the marginal densities

$$f(x,y)=(3/4)\,xy^2=f_1(x)\,f_2(y)=2\mathrm{x}\cdot(3/8)\,y^2 \text{ for } 0<x<1,\ 0<y<2 \quad (=0 \text{ else })$$

so yes they are independent.

e) Compute the (cumulative) joint distribution function as well as the marginal (cumulative) distribution functions and verify independence this way :

$$F(x,y)=\int_0^x\int_0^y (3/4)\,s\,t^2\,dt\,ds=(1/8)\,x^2\,y^3 \text{ for } 0<x<1,\ 0<y<2$$

$$F_1(x)=\int_0^x 2\mathrm{s}\,ds=x^2 \text{ for } 0<x<1 \text{ and } F_2(y)=\int_0^y (3/8)t^2\,dt=(1/8)\,y^3 \text{ for } 0<y<2$$

Since $F(x,y)=(1/8)\,x^2\,y^3=F_1(x)\,F_2(y)=x^2(1/8)\,y^3$ for $0<x<1,\ 0<y<2$

it would appear that the random variables are independent but one should really check the condition also holds when $1 < x$ so that $F(x,y)=(1/8)\,y^3=F_2(y)$ for $1<x, 0<y<2$ (= 1 when 2< y) etc

EXAMPLE 5 Joint uniform density : Particulate pollution in air samples collected from the smokestack of a coal fueled power plant decreases when a cleaning device is used. Let

X = amount of pollutant per sample when the cleaning device is not operating

Y = amount of pollutant per sample when the cleaning device is operating

It is observed that there is always at least twice as much particulate pollution without the cleaning device

as with it : $X>2\mathrm{Y}$ and the joint density is observed to be uniform over a triangular shaped region. Namely there is a positive constant *k* such that :

$$f_{X,Y}(x,y)=k \text{ for } 0\le x\le 2,\ 0\le y\le 1,\ 2\mathrm{y}\le x$$
$$f_{X,Y}(x,y)=0 \text{ else}$$

a) Find the value *k* that makes this a density :

Remarks on how to do a multiple integral : The region of integration in this problem is the 2-dimensional subset $A=\{(x,y): 0\le x\le 2,\ 0\le y\le 1,\ 2\mathrm{y}\le x\}\subset R^2$ a triangle shaped region in the plane. We can choose to do the integrals in different orders (x variable on the inside and y on the outside or visa versa), but once having fixed the outermost integration variable one always evaluates the innermost integral first, treating any outer variables in the integrand as fixed constants when performing the integral, then repeat the process. The limits of integration reflect this. With x as the outer variable the limits for y are determined by the conditions that for fixed x $0\le y\le x/2$. The outermost variable always goes between constant values in this case $0\le x\le 2$ and should not depend on the inner variables of integration. The limits of integration for inner integrals should either be constant or depend on outer variables of integration. The integral of the constant *k* here with respect to *y* gives *ky* which we evaluate between the limits 0 and *x*/2 to get *kx*/2. Being a constant the *k* factors outside the integrals.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy \;=\; 1 \;=\int_0^2\int_0^{x/2} k\,dy\,dx=k\int_0^2 x/2\,dx=(k\,x^2/4)_0^2=k \quad \text{gives } k=1$$

Thus the joint density is the constant 1 uniformly over the above triangular region and is 0 elsewhere.

Note we could have done the integral in the reverse order but the limits of integration
would change to reflect this. Then looking at the triangular region for fixed y we have $2y \le x \le 2$:

$$1 = \int_0^1 \int_{2y}^2 k\,dx\,dy = \int_0^1 k(2-2y)\,dy = k(2y - y^2)_0^1 = k \text{ gives } k = 1 \text{ as before} \quad .$$

b) Find the probability that the cleaning device will reduce the amount of pollutant by a factor of 3 or more : $P(X \ge 3Y) = P(Y \le X/3) = \int_0^2 \int_0^{x/3} 1\,dy\,dx = \int_0^2 x/3\,dx = 2/3$

c) Find the marginal densities for X and Y . Are X and Y independent random variables ? Explain.
The marginal density for X is

$$f_X(x) = \int_0^{x/2} f(x,y)\,dy = \int_0^{x/2} 1\,dy = x/2 \quad \text{for } 0 \le x \le 2 \quad (= 0 \text{ else })$$

The marginal density for Y is

$$f_Y(y) = \int_{2y}^2 f(x,y)\,dx = \int_{2y}^2 1\,dx = 2 - 2y \quad \text{for } 0 \le y \le 1 \;(= 0 \text{ else })$$

If X and Y were independent the joint density which is the constant 1 would factor as the product of the marginal densities for X and Y . But on the triangle shaped region where the joint density is 1

$$f(x,y) = 1 \ne f_X(x) f_Y(y) = (x/2)(2-2y) \quad \text{for } 0 \le x \le 2,\; 0 \le y \le x/2 \quad .$$

So X and Y are not independent.

d) Find the covariance of X and Y . Recall $cov[X,Y] = E[XY] - \mu_X \mu_Y$

$$\mu_X = E[X] = \int_0^2 x\, f_X(x)\,dx = \int_0^2 x^2/2\,dx = 4/3$$

$$\mu_Y = E[Y] = \int_0^1 y\, f_Y(y)\,dy = \int_0^1 y(2-2y)\,dy = (y^2 - 2y^3/3)_0^1 = 1/3$$

$$E[XY] = \int_0^2 \int_0^{x/2} xy\, f(x,y)\,dy\,dx = \int_0^2 \int_0^{x/2} xy\,dy\,dx = \int_0^2 x(y^2/2)_0^{x/2}\,dx = \int_0^2 x^3/8\,dx = (x^4/32)_0^2 = 1/2$$

This gives $cov[X,Y] = E[XY] - \mu_X \mu_Y = 1/2 - 4/9 = 1/18$.

e) Find the variance of X and Y. Then find the correlation of X and Y

$$E[X^2] = \int_0^2 x^2 f_X(x)\,dx = \int_0^2 x^3/2\,dx = (x^4/8)_0^2 = 2 \quad \text{so}$$

$$V[X] = E[X^2] - \mu_X^2 = 2 - (4/3)^2 = 2/9 \quad \text{is the variance of } X$$

$$E[Y^2] = \int_0^1 y^2 f_Y(y)\,dy = \int_0^1 y^2(2-2y)\,dy = (2y^3/3 - 2y^4/4)_0^1 = 1/6 \quad \text{so}$$

$$V[Y] = E[Y^2] - \mu_Y^2 = 1/6 - (1/3)^2 = 1/18 \quad \text{is the variance of } Y$$

To find the correlation we must divide the covariance 1/18 by the standard deviations of X and Y . But since these are just the square roots of the variances we get

$$\rho(X,Y) = cov[X,Y]/(\sigma_X \sigma_Y) = 1/18/(\sqrt{2/9}\sqrt{1/18}) = 1/2$$

EXAMPLE 6 Consider two continuous random variables X and Y representing the lifetime of two components in a system and with joint density

$$f(x,y) = 2e^{-x}e^{-2y} \text{ for } x > 0,\; y > 0$$

a) Find the marginal densities of the two lifetimes :

$$f_1(x)=f_X(x)=\int_0^\infty 2e^{-x}e^{-2y}\,dy=e^{-x} \text{ for } x>0$$ gives the marginal density for X

$$f_2(y)=f_Y(y)=\int_0^\infty 2e^{-x}e^{-2y}\,dx=2e^{-2y} \text{ for } y>0$$ gives the marginal density for Y

b) Are X and Y independent ?

The joint density $f(x,y)=2e^{-x}e^{-2y}=e^{-x}\cdot 2e^{-2y}=f_X(x)\cdot f_Y(y)$ for $x>0,\ y>0$ factors as the product of the marginal densities so yes, this show that X and Y are independent exponentially distributed random variables.

c) Compute the covariance of X and Y : in both the discrete and the continuous case we claim that the **covariance is 0 for independent random variables**. This example gives the general idea why :

$$cov[X,Y]=E[(X-\mu_X)(Y-\mu_Y)]=\int_0^\infty\int_0^\infty (x-\mu_X)(y-\mu_Y)2e^{-x}e^{-2y}\,dy\,dx$$

$$=\int_0^\infty (x-\mu_X)e^{-x}dx\int_0^\infty (y-\mu_Y)2e^{-2y}\,dy=E[X-\mu_X]E[Y-\mu_Y]=0$$

All that we have used is the fact that the joint density factors as the product of the marginal densities for independent random variables. (Similar reasoning holds in the discrete case)

Note : the **converse is not true** in general : the **covariance is 0 does not imply independence**.

EXAMPLE 7 For a simple example let X be a random variable with

$$P(X=0)=P(X=1)=P(X=-1)=1/3 \text{ and let } Y=1 \text{ if } X=0, \text{ and } Y=0 \text{ else}\ .$$

Then $XY=0$ so $E[XY]=0$. Also $E[X]=0$ so $Cov(X,Y)=E[XY]-E[X]E[Y]=0$.

However X and Y are clearly not independent since for example

$$P(Y=1)=P(X=0)=1/3=P(X=0 \text{ and } Y=1)\neq P(X=0)P(Y=1)=1/9$$

Expectation and Variance of a linear combination of continuous random variables :

For any constants a, b we have by standard properties of the integral and the order of integration

$$E[aX+bY]=\int_{-\infty}^\infty\int_{-\infty}^\infty (ax+by)f(x,y)\,dx\,dy=a\int_{-\infty}^\infty x\int_{-\infty}^\infty f(x,y)\,dy\,dx+b\int_{-\infty}^\infty y\int_{-\infty}^\infty f(x,y)\,dx\,dy$$

$$=a\int_{-\infty}^\infty x\,f_X(x)\,dx+b\int_{-\infty}^\infty y\,f_Y(y)\,dy=aE[X]+bE[Y]$$

For the above expected value property we do not need independence of X and Y
but for the variance calculation below we do assume independence of X and Y :

$$V[aX+bY]=E(a(X-\mu_X)+b(Y-\mu_Y))^2=a^2V[X]+b^2V[Y] \quad (X \text{ and } Y \text{ independent})$$

Here we have simply squared out the above expression and used the properties of expectation found above. The cross term vanishes since the covariance of X and Y is 0 using independence.

The above argument actually shows that

$$V[aX+bY]=E(a(X-\mu_X)+b(Y-\mu_Y))^2=a^2V[X]+2ab\,Cov(X,Y)+b^2V[Y]$$

in the general case for the variance of a linear combination **when X and Y are not independent.**

Since we showed the above in both the discrete and continuous case, repeating this process we find

Properties of expectation and variance : For random variables

$$X_i \text{ with mean } \mu_i=E[X_i] \text{ and variance } \sigma_i^2=V[X_i] \text{ for } i=1,2,\ldots,k$$

the linear combination $Y=a_1X_1+a_2X_2+\ldots+a_kX_k$ for any constants $a_1, a_2, \ \ldots\ , a_k$ has mean

$$E[Y]=E[\sum_{i=1}^k a_iX_i]=\sum_{i=1}^k a_iE[X_i] \quad \text{or} \quad \mu_Y=\sum_{i=1}^k a_i\mu_i$$

This just says expectation is a linear operation whether or not the variables in the sum are independent.

When the random variables are independent the variance of Y is given by

$$V[Y] = V[\sum_{i=1}^{k} a_i X_i] = \sum_{i=1}^{k} a_i^2 V[X_i] \quad \text{or} \quad \sigma_Y^2 = \sum_{i=1}^{k} a_i^2 \sigma_i^2$$

Note we have already seen the special case when k=2 with $X_2 = b$ a constant :

$$E[aX+b] = aE[X]+b \quad \text{and} \quad V[aX+b] = a^2 V[X]$$

If X and Y are independent then since the joint density (or pmf) factors for any functions h and g one has

$$E[h(X)\, g(Y)] = E[h(X)]\, E[g(Y)]$$

Conversely if the above property holds, the particular choice $h(x) = 1_{[-\infty,a]}(x)$; $g(y) = 1_{[-\infty,b]}(y)$ that h and g are indicator functions of intervals (h equal to 1 for x in the interval 0 else and similarly for g) shows that the joint cumulative distribution function factors into the product of the marginal cumulative distribution functions which implies independence , **so the above condition is equivalent to independence.**

EXAMPLE 8 If X has mean 1 and variance 3 and Y has mean 2 and variance 4 find

a) $E[\, 2X + 3\, Y + 4\,]$
$= 2E[X] + 3E[Y] + 4 = 2 + 6 + 4 = 12$

b) assuming X and Y are independent find the Variance $V[\, 2X + 3Y + 4]$
$= 2^2 V[X] + 3^2 V[Y] = 4\cdot 3 + 9\cdot 4 = 48$ (note the constant 4 doesn't contribute to the variance)

c) assuming independence find $V[2\text{XY}+3\text{X}+4]$

By the basic formula for variance this is $= E[(2\text{XY}+3\text{X}+4)^2] - (E[2\text{XY}+3\text{X}+4])^2$

Squaring out the first expression and using properties of expectation on the second the above equals

$$= E[4X^2Y^2 + 9X^2 + 16 + 12X^2Y + 16XY + 24X] - (2E[X]E[Y] + 3E[X] + 4)^2$$

$$= 4E[X^2]E[Y^2] + 9E[X^2] + 16 + 12E[X^2]E[Y] + 16E[X]E[Y] + 24E[X] - (2\cdot 1\cdot 2 + 3\cdot 1 + 4)^2$$

Noting that $E[X^2] = V[X] + (E[X])^2 = 3 + 1^2 = 4$ and $E[Y^2] = V[Y] + (E[Y])^2 = 4 + 2^2 = 8$ the above becomes $4\cdot 4\cdot 8 + 9\cdot 4 + 16 + 12\cdot 4\cdot 2 + 16\cdot 1\cdot 2 + 24\cdot 1 - 11^2$ = 211

Mean of the sample mean and sample variance : We have already looked at the case where we are sampling independently from an infinite population or sampling with replacement so that all the variables have the same mean $\mu = E[X_i]$ and same variance $\sigma^2 = V[X_i]$. Then we saw using the above properties that the sample mean has mean

$$E[\bar{x}] = E[\frac{S_n}{n}] = \frac{1}{n}\sum_{k=1}^{n} E[X_i] = \frac{1}{n} n\mu = \mu$$

and variance $V[\bar{x}] = V[\frac{S_n}{n}] = \left(\frac{1}{n}\right)^2 \sum_{k=1}^{n} V[X_i] = \left(\frac{1}{n}\right)^2 n\sigma^2 = \frac{\sigma^2}{n}$

Johnson works out the mean (expected value) of the sample variance using the above properties of the mean and variance of sums, by rewriting $(X_i - \bar{x}) = (X_i - \mu) + (\mu - \bar{x})$ He shows that it equals the actual population variance so s^2 is an unbiased estimator for σ^2 . That is :

The sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{x})^2$ has mean ; $E[s^2] = \sigma^2$

Linear combinations of normals : We have mentioned that the sum of two independent gamma distributed random variables with the same parameter β but with possibly different α parameters will still have a gamma distribution with the same parameter β and with the

A similar behavior holds for sums of independent normally distributed random variables. A linear combination or a **sum of independent normals is normal**. This is not hard to show but requires discussion of the **distribution of a sum** (the so-called **convolution** of the distributions) which we will not discuss here. When we showed that a standardized normal is normal N(0,1) we essentially showed that a constant times a normal random variable is still normal but we did not show this for general sums. If we believe the sum is normal then by the above properties of the expectation and the variance we know exactly the mean and variance of any linear combination of independent normals and knowing the result is normal we know exactly the **distribution of any linear combination of independent normals**,

For $X_i \sim N(\mu_i, \sigma_i^2)$ independent normals , the linear combination will be normal with distribution

$$Y = a_1 X_1 + a_2 X_2 + \ldots + a_k X_k \sim N(\mu_Y = \sum_i a_i \mu_i \ , \ \sigma_Y^2 = \sum_i a_i^2 \sigma_i^2)$$

EXAMPLE 9 Jill's bowling scores are normally distributed with mean 170 and standard deviation 20. Jack's are approximately normal with mean 160 and standard deviation 15. If Jack and Jill each bowl one game, assuming their scores are independent random variables, find the probability that :

a) Jack's score X is higher than Jill's score Y:

Note that $X - Y$ is normal with mean $E[X-Y] = E[X] - E[Y] = 160 - 170 = -10$ and variance

$$V[X-Y] = V[X] + (-1)^2 V[Y] = V[X] + V[Y] = \sigma_X^2 + \sigma_Y^2 = 15^2 + 20^2 = 25^2$$

Thus $P(X > Y) = P(X - Y > 0) = P(Z = \dfrac{X-Y-(-10)}{25} > \dfrac{0-(-10)}{25} = 2/5 = .4) = 1 - F_Z(.4)$

= 1 - .6554 = .3446.

b) The sum of their scores is above 350 :

$X+Y$ is normal with mean 160+170 = 330 and standard deviation 25 (same as for X-Y) so

$$P(X+Y > 350) = P(Z = \frac{X+Y-330}{25} > \frac{350-330}{25} = .8) = 1 - F_Z(.8)$$

= 1-.7881 = .2119

Normal scores plot : checking if the data is normal. Although the method is for simplicity illustrated for small samples (say n = 4 observations) in practice one wants 15 or 20 data observations. If the n observations are ordered from smallest to largest n data points will divide the real line into n+1 intervals. If the data comes from a standard normal distribution, the percentiles of the data should with high probability be approximately those corresponding to a standard normal distribution, with roughly 1/(n+1) being the probability under the normal curve of a normal value smaller than the smallest data value, or of lying in any one of the n+1 intervals determined by the ordered data. These idealized n **normal scores**

$$m_k \text{ such that } F(m_k) = \frac{k}{n+1} \ ; \ \ k = 1,2,\ldots,n$$

are easy to calculate using the cumulative distribution function F for the normal.

One then plots the pairs normal score value = x-coordinate with the actual value = y coordinate for the data. If the data is standard normal we expect to see this data plot lying close to a line with slope 1 (= 45 degrees) but for any other normal distribution the standardized z values which will be standard normal are related to the original normal data x by

$z = \dfrac{x-\mu}{\sigma}$ or equivalently $x = \sigma z + \mu$

so that we expect normal data will be close to linear (with slope σ). For n = 4 Johnson lists the normal scores

$$m_1 = z_{-.20} = -.84\,,\ m_2 = -z_{.40} = -.25\,,\ m_3 = z_{.40} = .25\,,\ m_4 = z_{.20} = .84$$

Recall we are using the ***z*-critical notation** where z_α is the value where $F(z_\alpha) = 1-\alpha$, that is the value which has area α to the right of it and $1-\alpha$ to the left of it under the standard normal curve.

Other authors use somewhat different definitions of the normal scores. For example Jay Devore has the *n* ordered data values lying close to the middle of *n* equal intervals so that for him the normal scores are determined by

$$m_k \text{ such that } F(m_k) = \frac{k-.5}{n}\,;\ \ k = 1,2,\ldots,n \quad \text{(different definition)}$$

For a large number of observations it doesn't make much difference which definition we use. In both cases one expects to see the normal plot of the data lying close to a straight line if the data comes from a normal distribution.

Transforming the data to near normality : If the normal plot of the data does not resemble a straight line, one can still hope that one can find (by trial and error or any other insight one might have) some function which when applied to the data gives a normal plot that is close (or at least closer) to linear. This is the method suggested by Johnson in section 5.12 of the text.

Johnson's method in section 5.12 is not the only way to proceed. Another important technique is to do a **probability plot** that is other than a normal one. That is for example if one has reason to believe the data comes from a Chi-squared distribution (with some fixed parameter) one can then determine the various percentiles that evenly divide that cumulative distribution (of that Chi-squared say rather than the standard normal distribution) into *n*+1 equal area segments (or some other variation such as Devore's) and then do this probability plot . If we get something close to a straight line we would then conclude the data comes from something close to a Chi-squared distribution and similarly for any other distribution that we suspect the data follows.

EXAMPLE 10 : Problem 5.95 of text.

a) For any 11 observations Johnson's method says we want to find the 11 values from table 3

$$m_k \text{ such that } F(m_k) = \frac{k}{12} = k(.083\bar{3})\,;\ \ k = 1,2,\ldots,11$$

or $\ -z_{.08\bar{3}}\,,\ -z_{.16\bar{6}}\,,\ -z_{.25}\,,\ -z_{.33\bar{3}}\,,\ -z_{.41\bar{6}}\,,\ z_{.5} = 0\,,\ z_{.41\bar{6}}\,,\ z_{.33\bar{3}}\,,\ z_{.25}\,,\ z_{.16\bar{6}}\,,\ z_{.08\bar{3}}$

using the z-critical notation. Thus by symmetry there are only 5 values we need to find to get all 11 values : -1.38 , -.97, -.67 , -.43 , -.21 , 0 , .21 , .43 , .67 , .97 , 1.38

For the neutrino data on p15 of the text to get the normal plot we first order the data as done on p36 of the text and then we plot the 11 points in the x-y plane :

(-1.38 , .021) , (-.97, .107), (-.67, .179), (-.43, .190), (-.21, .196) , (0, .283) , (.21, .580), (.43, .854), (.67, 1.18), (.97, 2.00) , (1.38, 7.30)

A plot of these points does not look linear but looks like the inter-arrival times are increasing exponentially when plotted against the normal scores which suggests taking the logarithm of the ordered data values. The log *X* values are -3.86, -2.23, -1.72, -1.66, -1.63, -1.26, -.54, -.16, .17, .69, 1.99 The plot of these values against the normal scores looks somewhat more linear suggesting that the data are closer to a log normal distribution.

Simulation of a distribution for continuous random variables :

If we wish to simulate independent identically distributed values of a continuous random variable *X* drawn from a distribution with density *f*(*x*) and cumulative distribution function *F*(*x*) then assume we can simulate independent copies of a uniform [0,1] random variable *U* obtaining a value *u*. Many computational statistics software packages such as the R platform will have built in commands to do this. If we don't have access to such software we can for example still simulate a uniform [0,1] to 3 place accuracy say by selecting 3 columns from a random digit table where we interpret the 3 digits in each row for the columns selected as representing 3 digits to the right of the decimal point thus producing random 3 decimal place numbers ranging from .000 to .999 (i.e. approximately uniform over the interval [0,1]). Then if we can solve the equation

$$F_X(x)=u$$

by inverting it for *x* corresponding to the value *u* obtained by simulating a uniform [0,1] random variable, the claim is that the the values x so obtained will simulate the random variable X with the desired distribution. This is intuitive since the cdf *F* increases from 0 to 1 and represents the probability = area under the density curve to the left of *x* , so the above is saying that we choose the values of x with uniform probability (equal likelihood) based on the desired distribution function.

To solve the above equation we need the function *F* to be 1-1 hence invertible. The cdf F is always increasing but will be 1-1 provided it strictly increases. This will be true where the density function is strictly positive. The claim is that for such a density

$$F_X(X)=U \quad \text{or equivalently} \quad X=F_X^{-1}(U) \quad .$$

That such a *U* is uniform is a consequence of the chain rule since to show that the density of *U* is the constant uniform density 1 on the interval [0,1] we have that the density

$$f_U(u)=\frac{d}{du}P(U=F_X(X)\le u) \quad \text{is the derivative of the cdf for } U \text{ which is given by}$$

$$F_U(u)=P(U=F_X(X)\le u)=P(X\le F_X^{-1}(u))=F_X(F_X^{-1}(u))=u \quad \text{for} \quad 0\le u\le 1$$

$$\text{and so} \quad f_U(u)=\frac{du}{du}=1 \quad \text{for} \quad 0\le u\le 1 \quad \text{as claimed}$$

The density of *U* is 0 outside the interval $0\le u\le 1$ since the cdf *F* can never go outside this interval.

The above method works well for distributions such as for exponential random variables with

$$F_X(x)=1-e^{-\lambda x} = u \Leftrightarrow x = -\frac{1}{\lambda}\ln(1-u)$$

where we know how to explicitly write down the cumulative distribution function .
Similarly for a random variable *X* with Weibull distribution since $Y=X^{\beta}$ is an exponential random variable we have

$$F_Y(y)=P(Y\le y)=1-e^{-\alpha y}=P(X^{\beta}\le y=x^{\beta})=P(X\le x)=F_X(x)$$

we find after substituting $y=x^{\beta}$ above the cdf of a Weibull is

$$F_X(x) = 1-e^{-\alpha x^{\beta}} = u \Leftrightarrow x = \left(-\frac{1}{\alpha}\ln(1-u)\right)^{1/\beta}$$

so the above method tells us how to simulate from a Weibull.

For normal random variables the preferred method is slightly different however. One draws $u_1,\ u_2$ from two independent uniform [0,1] random variables and then produces
two independent standard normal values (the *x* and *y* coordinates of a polar coordinate random point in the plane)

$$z_1=\sqrt{-2\ln(u_2)}\cos(2\pi u_1) \quad \text{and} \quad z_2=\sqrt{-2\ln(u_2)}\sin(2\pi u_1)$$

Note that the first uniform is used to select a uniform polar angle $\theta = 2\pi u_1$ between 0 and 2π while from the basic trig identity we see that in terms of the polar radius *r* one has

$$z_1^2 + z_2^2 = r^2 = -2\ln(u_2) \quad \text{or} \quad u_2 = e^{-r^2/2}$$ is inverted to get the radius *r*.

In deriving the constant $\sqrt{2\pi}$ associated with a standard normal density, we used the trick which in effect converted the joint density of two independent standard normals into its polar coordinate version which after integrating out the θ variable from 0 to 2π which cancels the factor of (1 / 2π) leaves us with the marginal for the *r* variable having marginal cumulative distribution function

$$F(r) = P(R = \sqrt{z_1^2 + z_2^2} \le r) = \int_0^r e^{-s^2/2} s\, ds = 1 - e^{-r^2/2}$$

Noting that if u_2 is uniform [0,1] so is $1 - u_2$, if we set the later equal to the above cdf for *F*(*r*) according to the above procedure, we see that this is equivalent to setting $u_2 = e^{-r^2/2}$ and solving for *r*. This shows that the above procedure really does produce two independent standard normal random variables.

EXAMPLE 11 To simulate two values from a normal distribution with mean & S.D.
$\mu = 20$ and $\sigma = 4$
we first select 3 columns of the random digits table 7 say columns 9, 10 and 11 and then select two rows say row 7 and row 8. This gives the two 3 digit numbers which we place to the right of the decimal point
which we interpret as two uniform [0,1] random values

$$u_1 = .846 \text{ and } u_2 = .053 \text{ so } \cos(2\pi u_1) = \cos(5.316) = .567 \text{ and } \sin(2\pi u_1) = \sin(5.316) = -.823$$

$$r = \sqrt{-2\ln(u_2)} = 2.4238 \text{ so } z_1 = r\cos(2\pi u_1) = 1.374 \text{ and } z_2 = r\sin(2\pi u_1) = -1.995$$

gives two simulated standard normal values. Now we have that

$z = \dfrac{x - \mu}{\sigma}$ or $x = \mu + \sigma z = 20 + 4z$ will give $N(20, 4^2)$ normal values with $\mu = 20$ and $\sigma = 4$.

So we get $x_1 = 20 + 4(1.374) = 25.496$ and $x_2 = 20 + 4(-1.995) = 12.02$ are our two simulated normal values.