WINTER 2017/18 TERM 2 STAT 302: ASSIGNMENT 3
Due: 2pm on Tuesday March 27, 2018

Question 1: We are studying a collection of molluscs living on a large beach in northern BC. The molluscs live near the high-tide line and it is known that their location is uniformly distributed with minimum -1 and maximum +2 metres from this line. Their daily energy intake (in kilocalories) turns out to be 1 plus 25% of their squared location.

(a) What is the cumulative distribution function of the daily energy intake?

Let X denote the location of the molluscs. $X \sim U(-1, 2)$
Let Y denote the daily energy intake, $Y = 1 + 25\%X^2$

$$f(x) = \begin{cases} 1/3, & x \in [-1, 2] \\ 0, & otherwise \end{cases}$$

$$Pr(X \leqslant x) = F(x) = \begin{cases} 0, & x < -1 \\ \frac{x+1}{3}, & x \in [-1, 2] \\ 1, & x > 2 \end{cases}$$

$$Pr(Y \leqslant y) = Pr(1+0.25X^2 \leqslant y) = Pr(X^2 \leqslant 4(y-1)) = Pr(X \geqslant -2\sqrt{y-1}, X \leqslant 2\sqrt{y-1})$$

Compute the domain of y:

$$-1 \leqslant -2\sqrt{y-1} \leqslant 0 \implies 1 \leqslant y \leqslant \frac{5}{4}$$

$$0 \leqslant 2\sqrt{y-1} \leqslant 2 \implies 1 \leqslant y \leqslant 2$$

Therefore, when $1 \leqslant y \leqslant \frac{5}{4}, F(y) = \frac{2\sqrt{y-1}+1}{3} - \frac{-2\sqrt{y-1}+1}{3} = \frac{4\sqrt{y-1}}{3}$
So the cumulative distribution function of daily energy intake is:

$$F(y) = \begin{cases} 0, & y < 1 \\ \frac{4\sqrt{y-1}}{3}, & 1 \leqslant y \leqslant \frac{5}{4} \\ \frac{2\sqrt{y-1}+1}{3}, & \frac{5}{4} \leqslant y \leqslant 2 \\ 1, & y > 2 \end{cases}$$

(b) What is the probability density function of the daily energy intake?
$(\frac{2\sqrt{y-1}+1}{3})' = \frac{2}{3} \cdot \frac{1}{2} \frac{1}{\sqrt{y-1}} = \frac{1}{3\sqrt{y-1}}, 1 \leqslant y \leqslant \frac{5}{4}$
$(\frac{4\sqrt{y-1}}{3})' = \frac{4}{3} \cdot \frac{1}{2} \frac{1}{\sqrt{y-1}} = \frac{2}{3\sqrt{y-1}}, \frac{5}{4} \leqslant y \leqslant 2$

$$f(y) = F'(y) = \begin{cases} 0, & y < 1 \ or \ y > 2 \\ \frac{2}{3\sqrt{y-1}}, & 1 \leqslant y \leqslant \frac{5}{4} \\ \frac{1}{3\sqrt{y-1}}, & \frac{5}{4} \leqslant y \leqslant 2 \end{cases}$$

(c) What is the expected daily energy intake?

$$\begin{aligned}
\mathbb{E}(Y) &= \int_{-\infty}^{\infty} yf(y)dy = \int_{1}^{5/4} \frac{2y}{3\sqrt{y-1}}dy + \int_{5/4}^{2} \frac{y}{3\sqrt{y-1}}dy \\
&= \int_{1}^{5/4} \frac{2}{3}\sqrt{y-1}dy + \int_{1}^{5/4} \frac{2}{3\sqrt{y-1}}dy + \int_{5/4}^{2} \frac{1}{3}\sqrt{y-1}dy + \int_{5/4}^{2} \frac{1}{3\sqrt{y-1}}dy \\
&= \frac{4}{9}(y-1)^{3/2}|_{1}^{5/4} + \frac{4}{3}\sqrt{y-1}|_{1}^{5/4} + \frac{2}{9}(y-1)^{3/2}|_{5/4}^{2} + \frac{2}{3}\sqrt{y-1}|_{5/4}^{2} \\
&= \frac{4}{9} \times \frac{1}{8} + \frac{4}{3} \times \frac{1}{2} + \frac{2}{9} \times \frac{7}{8} + \frac{2}{3} \times \frac{1}{2} \\
&= \frac{5}{4}
\end{aligned}$$

Question 2: Let X and Y be two independent Bernoulli(0.5) random variables and define $U = X + Y$ and $V = X - Y$.

(a) Find the joint and marginal probability mass functions for U and V. [It is sufficient to construct a table to describe these mass functions.]

The joint probability mass function is:

Pr(U=u, V= v)	U=0	U=1	U=2
V=-1	0	1/4	0
V=0	1/4	0	1/4
V=1	0	1/4	0

The marginal probability function of U is:

U=u	U=0	U=1	U=2
$P_U(u)$	1/4	1/2	1/4

The marginal probability function of V is:

V=v	V=-1	V=0	V=1
$P_V(v)$	1/4	1/2	1/4

(b) Are U and V independent? Why or why not?

No. U and V are independent if and only if $p(u, v) = p_U(u) \cdot p_V(v)$.

For example, $p(u = 1, v = 1) = 1/4$ from the joint probability mass function, but $p_U(u = 1) = 1/2$ and $p_V(v = 1) = 1/4$ from the marginal probability functions; therefore $p_U(u = 1) \cdot p_V(v = 1) = 1/8$, so $p(u = 1, v = 1) \neq p_U(u = 1) \cdot p_V(v = 1)$.

By the counter example above, we can conclude that $p(u, v) = p_U(u) \cdot p_V(v)$ does not hold for the domain, so U and V are not independent.

(c) Find the conditional probability mass functions $p_{U|V=v}(u)$ and $p_{V|U=u}(v)$. [Again, you can construct a table to describe these mass functions.]

The conditional probability mass function $p_{U|V=v}(u)$

$Pr(U \mid V=v)$	$U=0$	$U=1$	$U=2$
fix $V=-1$	0	1	0
fix $V=0$	1/2	0	1/2
fix $V=1$	0	1	0

The conditional probability mass function $p_{V|U=u}(v)$

$Pr(V \mid U=u)$	$V=-1$	$V=0$	$V=1$
fix $U=0$	0	1	0
fix $U=1$	1/2	0	1/2
fix $U=2$	0	1	0

Question 3: This question will provide an intriguing contrast to Question 2. Recall that if we have a continuous random variable X defined by a pdf $f_X(x)$, and we define a new random variable $Y = g(X)$ where g is a bijective (i.e. one-to-one) transformation, then the inverse of g is well-defined everywhere, g^{-1}, and the density of Y is given by

$$f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right|.$$

We can generalize this to the bivariate setting as follows. Suppose X_1 and X_2 are continuous random variables with joint pdf $f_{X_1,X_2}(x_1,x_2)$ and suppose that both $u = g_1(X_1,X_2)$ and $v = g_2(X_1,X_2)$ are bijective (i.e. one-to-one) transformations with inverses $g_1^{-1}(u,v)$ and $g_2^{-1}(u,v)$. If these inverse functions have continuous partial derivatives and nonzero *Jacobian*

$$J = \det \begin{bmatrix} \frac{\partial g_1^{-1}}{\partial u} & \frac{\partial g_1^{-1}}{\partial v} \\ \frac{\partial g_2^{-1}}{\partial u} & \frac{\partial g_2^{-1}}{\partial v} \end{bmatrix} = \frac{\partial g_1^{-1}}{\partial u}\frac{\partial g_2^{-1}}{\partial v} - \frac{\partial g_2^{-1}}{\partial u}\frac{\partial g_1^{-1}}{\partial v} \neq 0,$$

then the joint density of u and v is

$$f_{u,v}(u,v) = f_{X_1,X_2}\left(g_1^{-1}(u,v), g_2^{-1}(u,v)\right)|J|,$$

where $|J|$ is the absolute value of the Jacobian.

(a) Let X_1 and X_2 be independent standard normal random variables. Write down the joint probability density function of X_1 and X_2. Moreover, compute $\Pr(X_1 < 1, X_2 < 1)$.

$X_1 \sim N(0,1)$, and $X_2 \sim N(0,1)$
$f(x_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_1^2}{2}}$
$f(x_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x_2^2}{2}}$
The joint probability density function is:
$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi} e^{-\frac{x_1^2+x_2^2}{2}}$

$Pr(X_1 < 1, X_2 < 1) = Pr(X_1 < 1) \cdot Pr(X_2 < 1) = 0.8413^2 = 0.70778569$

(b) Define the transformations $u = g_1(x_1,x_2) = x_1 + x_2$ and $v = g_2(x_1,x_2) = x_1 - x_2$. Find the inverse functions $g_1^{-1}(u,v)$ and $g_2^{-1}(u,v)$ and compute the Jacobian of this bivariate transformation of variables.

$g_1(u,v) = u + v = 2x_1 \implies x_1 = \frac{u+v}{2}$
$g_2(u,v) = u - v = 2x_2 \implies x_2 = \frac{u-v}{2}$

Therefore, $g_1^{-1}(u,v) = \frac{u+v}{2}$ and $g_2^{-1}(u,v) = \frac{u-v}{2}$

$$J = \det \begin{bmatrix} \frac{\partial g_1^{-1}}{\partial u} & \frac{\partial g_1^{-1}}{\partial v} \\ \frac{\partial g_2^{-1}}{\partial u} & \frac{\partial g_2^{-1}}{\partial v} \end{bmatrix} = \frac{\partial \frac{u+v}{2}}{\partial u}\frac{\partial \frac{u-v}{2}}{\partial v} - \frac{\partial \frac{u-v}{2}}{\partial u}\frac{\partial \frac{u+v}{2}}{\partial v} = -\frac{1}{4} - \frac{1}{4} = -\frac{1}{2},$$

(c) Write down the joint pdf of $U = X_1 + X_2$ and $V = X_1 - X_2$ and show that this density separates over variables; i.e. show $f_{U,V}(u,v) = a(u)b(v)$ for some real functions $a(u)$ and $b(v)$. Recall from class that this implies that U and V are actually *independent.*

The joint pdf of U and V is:
$f_{u,v}(u,v) = f_{X_1,X_2}\left(g_1^{-1}(u,v), g_2^{-1}(u,v)\right)|J| = f_{X_1,X_2}\left(\frac{u+v}{2}, \frac{u-v}{2}\right)\frac{1}{2} = \frac{1}{4\pi}e^{-\frac{(u+v)^2+(u-v)^2}{8}}$
$= \frac{1}{4\pi}e^{-\frac{u^2+v^2}{4}} = \frac{1}{2\sqrt{\pi}}e^{-\frac{u^2}{4}} \cdot \frac{1}{2\sqrt{\pi}}e^{-\frac{v^2}{4}} = a(u)\cdot b(v)$

Therefore, we have found $a(u) = \frac{1}{2\sqrt{\pi}}e^{-\frac{u^2}{4}}$, and $b(v) = \frac{1}{2\sqrt{\pi}}e^{-\frac{v^2}{4}}$
So the density separates over variables.

(d) Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. If $n = 2$, show that the sample mean, $\bar{X}$, and the sample variance, S_X^2, are independent random variables. [Hint: first write $\bar{X}$ and S_X^2 in terms of U and V as above, remembering that U and V are linear combinations of *standard* normal random variables.]

Standarize X_1 and X_2: $X_1' = \frac{X_1-\mu}{\sigma}$, $X_2' = \frac{X_2-\mu}{\sigma}$
Take the definition in (c), $U = X_1' + X_2', V = X_1' - X_2'$
The sample mean

$$\bar{X} = \frac{X_1 + X_2}{2} = \frac{(\sigma X_1' + \mu) + (\sigma X_2' + \mu)}{2} = \frac{\sigma}{2}U + \mu = f(u)$$

The sample variance

$$S_X^2 = \frac{1}{2-1}((X_1-\bar{X})^2+(X_2-\bar{X})^2) = \frac{(X_1 - X_2)^2}{2} = \frac{(\sigma X_1' + \mu - \sigma X_2' - \mu)^2}{2} = \frac{\sigma^2 V^2}{2} = g(v)$$

In question (c) we have proved that $f_{u,v}(u,v) = a(u)\cdot b(v)$, which means U and V are independent random variables. $f_{u,v}(f(u), g(v)) = a(f(u))\cdot b(g(v))$

$$f(\bar{X}, S_X^2) = f_{u,v}(\frac{\sigma}{2}U + \mu, \frac{\sigma^2 V^2}{2}) = a(f(u))\cdot b(g(v))$$

Therefore, $\bar{X}$ and S_X^2 are independent of each other.

Question 4: Let W be a Gamma random variable with parameters $(\alpha = 2,\ \lambda = 2)$. Conditional on the value $W = w$, X is an exponential random variable with rate parameter w.

(a) What is the conditional density function for X given $W = w$? Be sure to indicate any restrictions on the values of x and w. If $W = 2$, what is the probability that $X \leqslant 2$?

$W \sim Gamma(2, 2)$
$\Gamma(2) = (2-1)! = 1$
$f(w) = \begin{cases} 4e^{-2w}w, & w \geqslant 0 \\ 0, & otherwise \end{cases}$

$X \sim Exp(w)$, for a fixed w
$f_{X|W=w}(x) = \begin{cases} we^{-wx}, & x \geqslant 0 \\ 0, & otherwise \end{cases}$

When $W = 2$, $f_{X|W=2}(x) = \begin{cases} 2e^{-2x}, & x \geqslant 0 \\ 0, & otherwise \end{cases}$

The probability $Pr(X \leqslant 2) = \int_0^2 2e^{-2x}dx = 1 - e^{-4}$

(b) What is the probability that W is greater than its expected value? Do *not* use an online applet to find this probability; calculate by hand.

$\mathbb{E}(W) = \frac{\alpha}{\lambda} = 1$
$Pr(W > 1) = 1 - Pr(W \leqslant 1) = 1 - \int_0^1 4e^{-2w}wdw = 1 + 2\int_0^1 -2e^{-2w}wdw$
$= 1 + 2(e^{-2w}w|_0^1 - \int_0^1 e^{-2w}dw)$
$= 1 + 2(e^{-2} + \frac{1}{2}(e^{-2} - 1))$
$= 3e^{-2}$

(c) Show that the conditional distribution of W given $X = 2$ is Gamma distributed with parameters $(\alpha = 3,\ \lambda = 4)$. [Recall that $\Gamma(\alpha) = (\alpha - 1)!$, and use the fact that $f_X(2) = 1/8$.]

(1) $f_{(W|X=2)}(w) = \frac{f(X=2,\ W=w)}{f_X(2)} = \frac{f(X=2|W=w)f(w)}{f_X(2)} = \frac{we^{-2w} \cdot 4e^{-2w}w}{1/8} = 32e^{-4w}w^2, w \geqslant 0$
$f_{(W|X=2)}(w) = \begin{cases} 32e^{-4w}w^2, & w \geqslant 0 \\ 0, & otherwise \end{cases}$

(2) The pdf for Gamma distribution with $\alpha = 3,\ \lambda = 4$ is
$f(w)_{(\alpha=3,\ \lambda=4)} = \begin{cases} \frac{4e^{-4w}(4w)^2}{\Gamma(3)} = 32e^{-4w}w^2, & w \geqslant 0 \\ 0, & otherwise \end{cases}$

(1) (2) $\implies f_{(W|X=2)}(w) = f(w)_{(\alpha=3,\ \lambda=4)}$. Therefore, the conditional distribution of W given X =2 is Gamma distributed with parameters $\alpha = 3,\ \lambda = 4$.

Question 5: Let X be the amount of time that a student spends walking from the Earth Sciences Building to the Mathematics Building, and let Y be the amount of time that a student spends walking from the Earth Sciences Building to the Pharmaceutical Sciences Building. Suppose that the joint density of X and Y is given by the following function

$$f_{X,Y}(x, y) = \begin{cases} \frac{kx}{y} & \text{if } 0 < x < y < 10 \\ 0 & \text{otherwise} \end{cases}$$

for some fixed constant k.

(a) What value of k makes $f_{X,Y}(x, y)$ an honest probability density function?

$$\int_0^{10}\int_0^{y}\frac{kx}{y}dxdy = \int_0^{10}\frac{k}{y}\frac{1}{2}y^2dy = \frac{k}{4}y^2|_0^{10} = 25k = 1$$
$$\implies k = \frac{1}{25}$$

(b) Find the marginal probability density function for X. [Don't forget to specify the support of the function!]

$$f_X(x) = \int_x^{10}\frac{kx}{y}dy = kxln(y)|_x^{10} = kx(ln(10) - ln(x)), 0 < x < 10$$

Therefore,
$$f_X(x) = \begin{cases} \frac{1}{25}x(ln(10) - ln(x)), & 0 < x < 10 \\ 0, & otherwise \end{cases}$$

(c) Find the probability $\Pr(Y + X \leqslant 10)$.
$Y + X \leqslant 10 \implies Y \leqslant -X + 10$

$$Pr(Y + X \leqslant 10)$$
$$= \int_0^{5}\int_0^{y}\frac{kx}{y}dxdy + \int_5^{10}\int_0^{10-y}\frac{kx}{y}dxdy$$
$$= \frac{k}{2}\int_0^{5}ydy + \frac{k}{2}\int_5^{10}\frac{(10-y)^2}{y}dy$$
$$= \frac{1}{50}(\frac{25}{2} + \int_5^{10}(y - 20 + \frac{100}{y})dy)$$
$$= \frac{1}{50}(\frac{25}{2} + \frac{75}{2} - 100 + 100(ln(10) - ln(5)))$$
$$= 2(ln(10) - ln(5)) - 1$$
$$= 0.38629436$$