Math 302, PSET 4

(1) Let U_1, U_2 be independent uniform random variables on $(0,1)$, and let $X = |U_1 - U_2|$.
  (a) Compute $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$.
  (b) Compute $\mathrm{Corr}(U_1, X)$.
  (c) Determine a formula for the conditional probability density $f_{U_1|X}(u|x)$ of U_1 given X.

Solution: (a) $\mathbb{E}[X] = \int_0^1 \int_0^1 |x-y|\,dx\,dy = 1/3; \mathbb{E}[X^2] = \int_0^1 \int_0^1 |x-y|^2\,dx\,dy = 1/6.$

(b) The key computation is $\mathbb{E}[U_1 X] = \int_0^1 \int_0^1 x|x-y|\,dx\,dy = 1/6$. Thus $\mathrm{Cov}(U_1, X) = \mathbb{E}[U_1 X] - \mathbb{E}[U_1]\mathbb{E}[X] = 1/6 - (1/2)(1/3) = 0 = \mathrm{Corr}(U_1, X)$.

(c) Note that the PDF of X is $f_X(s) = 2(1-s)$ for $s \in (0,1)$. This can be obtained by observing that the set $\{(x,y) \in [0,1]^2 : |x-y| \le s\}$ is the polygon with vertices $(0,0), (s,0), (1,1-s), (1,1), (1-s,1)$, and $(0,s)$, which has area $(1-s)^2 = F_X(s)$. For the joint PDF of X and U_1, the relevant probability is $\mathbb{P}(U_1 \le x, X \le s)$, which is given by the area of the same polygon, but truncated by the vertical line at x. The area inside the polygon and to the left of that line is given by the formula

$$F_{U_1,X}(x,s) = (\mathbb{P}(U_1 \le x, X \le s) = \begin{cases} \frac{1}{2}x^2 + sx, & x \in (0,s) \\ 2sx - \frac{1}{2}s^2, & x \in (s, 1-s) \\ 1 - \frac{1}{2}(1-x)^2 - s(1-x), & x \in (1-s, 1). \end{cases}$$

The densities f_X and $f_{U_1,X}$ can be obtained by taking derivatives.

(2) Let X be a Poisson(1) random variable, and let Y be the random variable distributed as Uniform$(0, X)$.
  (a) Compute $\mathbb{E}Y$.
  (b) Compute $\mathrm{Corr}(X, Y)$.
  (c) Determine a formula for the conditional probability density $f_{X|Y}(x|y)$ of X given Y.
  (d) What is the distribution of the conditional expectation $\mathbb{E}[X|Y]$? What about $\mathbb{E}[Y|X]$?

Solution: (a) Condition on X, using the fact that $\mathbb{E}[Y|X=0] = 0$:

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]] = \sum_{k=1}^{\infty} \mathbb{E}[Y|X=k]\mathbb{P}(X=k) = \sum_{k\ge1} \frac{k}{2} e^{-1}\frac{1}{k!} = \frac{1}{2}.$$

(b) By the same method as in (a),

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[XY|X]] = \sum_{k\ge1} \mathbb{E}[kY|X=k]\mathbb{P}(X=k) = \sum_{k\ge1} \frac{k^2}{2}\mathbb{P}(X=k) = 1.$$

Note that $\mathrm{Var}(X) = 1$, and since $\mathrm{Var}(\mathrm{Unif}(0,k)) = k^2/3$,

$$\mathbb{E}[Y^2] = \mathbb{E}[\mathbb{E}[Y^2|X]] = \sum_{k\ge1} \mathbb{E}[Y^2|X=k]\mathbb{P}(X=k) = \sum_{k\ge1} \frac{k^2}{3}\mathbb{P}(X=k) = 2/3.$$

Thus $\mathrm{Corr}(X,Y) = \frac{1-(1/2)\cdot 1}{\sqrt{1}\sqrt{2/3}} = \sqrt{6}/4$.

(c) The joint PDF is $f_{X,Y}(k,y) = \mathbb{P}(X=k, Y=y) = \frac{1}{e}\frac{1}{k!}\frac{1}{k}1\{0 \le y \le k\}$ for any integer $k \ge 1$ and any real number $y \ge 0$, while $\mathbb{P}(X=0, Y=0) = \frac{1}{e}$. (Note that this is a joint PDF of one discrete and one continuous r.v.!) The marginal distribution of Y is given by

$$f_Y(y) = \begin{cases} \frac{1}{e}, & y = 0 \\ \sum_{k=1}^{\infty} f_{X,Y}(k,y) = \sum_{k=\lceil y \rceil}^{\infty} \frac{1}{e}\frac{1}{kk!}, & y > 0 \end{cases}$$

where $\lceil \cdot \rceil$ is the 'ceiling' function, the smallest integer larger than $\cdot$.

(d) Given X, Y is uniform on $(0, X)$, so $\mathbb{E}[Y|X] = X/2$. This can be verified directly by computing the function $g(k) = \int_0^\infty y\mathbb{P}(Y=y|X=k)\,dy = k/2$. The distribution of $\mathbb{E}[X|Y]$ is given by the usual conditional formula:

$$\mathbb{E}[X|Y] = h(Y), \text{ where } h(y) = \sum_{k \ge 1} x f_{X|Y}(k,y) = \frac{1-e^{-1}}{f_Y(y)}.$$

(3) (Anderson, 4.16) Choose 500 numbers uniformly from the interval $[1.5, 4.8]$.

(a) Approximate the probability that less than 65 of the numbers start with the digit 1.

(b) Approximate the probability of the event that more than 160 of the numbers start with the number 3.

Solution:

(a) Let p be the probability that a sampled number starts with the digit 1, so $p = 0.5/(4.8-1.5) = 0.1515$. Let $n = 500$, and let X denote the total number of numbers selected which start with the digit 1, so $X \sim \text{Binomial}(n,p)$. We want to approximate the distribution of X.
Considering our options; by Theorem 4.20 (Anderson), the error of the Poisson Approximation is unacceptable in this problem. However, $np(1-p) > 10$ so we will use a Normal Approximation (Anderson p.g 139 for details).
Let $Z \sim N(0,1)$. Then,

$$\mathbb{P}(X < 65) \approx \mathbb{P}(Z \le \frac{64.5 - np}{\sqrt{np(1-p)}}) = \Phi(-1.404) \approx 0.0801$$

(b) We use the same method as in (a). In this case, $p = 1/(4.8-1.5) = 0.3030$.

$$\begin{aligned}\mathbb{P}(X > 160) &= 1 - \mathbb{P}(X \le 160) \approx 1 - \mathbb{P}(Z \le \frac{160.5 - np}{\sqrt{(np(1-p))}}) \\ &\approx 1 - \Phi(0.875) \approx 0.1908\end{aligned}$$

(4) Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli(1/2) random variables, i.e. a sequence of n coin flips.

(a) Let T_n be the number of indices $0 \le i \le n-2$ where X_i, X_{i+1} and X_{i+2} are all 1. Find $\mathbb{E}T$ and $\mathrm{Var}(T)$.

(b) Fix $n = 5$. Describe the distribution of the conditional expectation $\mathbb{E}[T|X_3]$ in terms of the X_i's.

Solution:

(a) Let $T^k := 1_{\{X_k=1,X_{k+1}=1,X_{k+2}=1\}}$. Notice that $E[T^k] = \mathbb{P}(X_k = 1, X_{k+1} = 1, X_{k+2} = 1) = 1/8$ for all k (By independence of the flips). Also, $T = T_1 + ... + T_{n-2}$, so by the linearity of expectation

$$E[T] = \sum_{k=0}^{n-2} E(T^k) = \frac{1}{8}(n-2)$$

We have $(E[T])^2$ from the above calculation, so we just need to compute $E[T^2]$.

$$E[T^2] = E[\sum_{j=1}^{n-2}\sum_{i=1}^{n-2} T^i T^j] = \sum_{j=1}^{n-2} E[(T^j)^2] + \sum_{i\neq j}^{n-2} E[T^i T^j]$$

Since the T^i's are indicators, notice that the first sum is simply $E[T]$. To compute the second sum, we consider 3 cases; 1) if $|i-j| > 2$ then the indicators T^i and T^j do not share any flip and so $E[T^iT^j] = 1/2^6$. 2) if $|i-j| = 2$ the indicators share one flip so $E[T^iT^j] = 1/2^5$. 3) if $|i-j| = 1$ the indicators share two flips so $E[T^iT^j] = 1/2^4$. Therefore

$$\begin{aligned}
\sum_{i\neq j}^{n-2} E[T^iT^j] &= \sum_{i\neq j,|i-j|>2}^{n-2} E[T^iT^j] + \sum_{i\neq j,|i-j|=2}^{n-2} E[T^iT^j] + \sum_{i\neq j,|i-j|=1}^{n-2} E[T^iT^j] \\
&= \frac{1}{2^6}(n^2 - n - 2(n-4) - 2(n-3)) + \frac{2}{2^5}(n-4) + \frac{2}{2^4}(n-3) \\
&= \frac{1}{2^6}(n^2 - 5n - 14) + \frac{1}{2^4}(n-4) + \frac{1}{2^3}(n-3)
\end{aligned}$$

So $E(T^2) = \frac{1}{2^3}(n-2) + \frac{1}{2^6}(n^2-5n-14) + \frac{1}{2^4}(n-4) + \frac{1}{2^3}(n-3)$ and finally $Var(T) = \frac{1}{2^3}(n-2) + \frac{1}{2^6}(n^2-5n-14) + \frac{1}{2^4}(n-4) + \frac{1}{2^3}(n-3) - \frac{1}{2^6}(n-2)^2$

(b) By the linearity of conditional expectation

$$E[T_5|X_3] = \sum_{k=1}^{3} E[T_5^k|X_3]$$

On the event $\{X_3 = 0\}$ then each T_5^k is zero since every 3 flip sequence within 5 flips includes the realization of X_3. Thus $E[T_5^k|X_3 = 0] = 0$ for each k, giving $E[T_5|X_3] = 0$. $\mathbb{P}(T_5^k = 1|X_3 = 1) = 1/2^2$. Therefore $\mathbb{E}[T_5|X_3] = 3X_3/4$.

(5) Let Z_1, Z_2 be independent N(0,1) random variables. Identify the distribution of $Z_1 + Z_2$ (it should be familiar) in two different ways:
 (a) Using the convolution formula.
 (b) Using moment generating functions.

Solution:

(a) To reduce subscripts let $X = Z_1$, $Y = Z_2$, $Z = X + Y$ By the convolution formula

$$\begin{aligned}
f_z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{\frac{-1}{2}z^2} e^{-x^2} e^{xz} dx \\
&= \frac{1}{2\pi} e^{\frac{-1}{2}z^2} e^{\frac{1}{4}z^2} \int_{-\infty}^{\infty} e^{-(x-\frac{z}{2})^2} dx \\
&= \frac{1}{2\sqrt{\pi}} e^{-\frac{z^2}{4}}
\end{aligned}$$

f_z is the pdf of a $N(0,2)$ RV. Therefore $Z \sim N(0,2)$

(b) By independence

$$\begin{aligned} M_Z(t) &= E[e^{t(X+Y)}] \\ &= E[e^{tX}]E[e^{tY}] \\ &= e^{t^2} \end{aligned}$$

The MGF of $N(0,2)$. The MGF of a RV determines it's distribution, hence $Z \sim N(0,2)$

(6) Let Z_1, Z_2 be independent N$(0,1)$ random variables. Identify the distribution of $Z_1^2 + Z_2^2$ (it should be familiar) in two different ways:
(a) Using the convolution formula.
(b) Using moment generating functions.

Solution:
(a) To use the convolution formula, we first need to work out $f_{X^2}(x)$. Assume that $x > 0$ then

$$\mathbb{P}(X^2 \le x) = \mathbb{P}(|X| \le \sqrt{x}) = \mathbb{P}(X \le \sqrt{x}) - \mathbb{P}(X \le -\sqrt{x})$$

That is, $F_{X^2}(x) = F_X(\sqrt{x}) - F_X(-\sqrt{x})$. Differentiating both sides gives

$$f_{X^2}(x) = \frac{1}{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} \quad x > 0$$

($f_{X^2}(z) = 0$ if $x \le 0$) So

$$\begin{aligned} f_z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx \\ &= \frac{1}{2\pi} \int_0^z e^{-x/2} e^{-(z-x)/2} \frac{1}{\sqrt{x}} \frac{1}{\sqrt{z-x}} dx \\ &= \frac{1}{2} e^{-\frac{z}{2}} \text{ (Not an easy integral!)} \end{aligned}$$

The PDF of $Exp(1/2)$. Therefore $Z \sim Exp(1/2)$
(b) Let $g(x) = e^{tx^2}$. Then $M_{X^2}(t) = E[e^{tX^2}] = E(g(X))$

$$\begin{aligned} M_{X^2}(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx^2} e^{-\frac{1}{2}x^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}(1-2t)} dx \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{1-2t}} \int_{-\infty}^{\infty} e^{\frac{-u^2}{2}} du \\ &= \frac{1}{\sqrt{1-2t}} \quad (t < 1/2) \end{aligned}$$

By independence $M_{X^2+Y^2}(t) = M_{X^2}(t) M_{Y^2}(t)$. So $M_Z(t) = 1/(1-2t)$, the MGF of $Exp(1/2)$. Therefore $Z \sim Exp(1/2)$

(7) Give an example of two jointly continuous random variables X, Y satisfying:
- X and Y are not independent
- X and Y do not have the same marginal distribution
- $\text{Cov}(X, Y) = 0$.

The random variables X and U_1 from problem 1 on this PSET are such an example. That they are not indendent is not too hard to see: for example, the event $U_1 \in (.4, .6)$ and $X \in (.9, 1)$ has probability 0, but $\mathbb{P}(U \in (.4, .6) = .2$ and $\mathbb{P}(X \in (.9, 1)) > 0$.

(8) Suppose X is a random variable with moment generating function $M_X(t) = \frac{1}{4}e^{-t} + \frac{1}{4} + \frac{1}{2}e^{4t}$.
  (a) Find the mean and variance of X by differentiating M.
  (b) Find the PMF of X, and use it to check your answers from part (a).

Solution:

(a) $E[X] = M'_X(0) = [-\frac{1}{4}e^{-t} + 2e^{4t}]_{t=0} = 7/4$. $E[X^2] = M''_X(0) = [\frac{1}{4}e^{-t} + 8e^{4t}]_{t=0} = 33/4$. So $Var(X) = 33/4 - 49/16 = 11/2$

(b) $M_X(t) = E(g(X)) =$ where $g(x) = e^{tx} = \frac{1}{4}e^{-t} + \frac{1}{4} + \frac{1}{2}e^{4t}$. So it follows that $\mathbb{P}(X = -1) = 1/4$, $\mathbb{P}(X = 0) = 1/4$, $\mathbb{P}(X = 4) = 1/2$. Thus, $E[X] = (-1 * \frac{1}{4}) + (0 * \frac{1}{4}) + (4 * \frac{1}{2})$ $= 7/4$ and $E[X^2] = (1 * \frac{1}{4}) + (0 * \frac{1}{4}) + (16 * \frac{1}{2}) = 33/4$

(9) (Anderson, 8.12) Let Z be Gamma$(2, \lambda)$ distributed for some $\lambda > 0$, i.e.

$$f_Z(z) = \begin{cases} \lambda^2 z e^{-\lambda z}, & z \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

  (a) Find the moment generating function of Z.
  (b) Let X, Y be independent Exponential(λ) random variables. Show that $X + Y$ has the same distribution as Z.

Solution:

(a)

$$\begin{aligned} M_Z(t) &= \int_{-\infty}^{\infty} e^{xt} f_z(x)dx \\ &= \lambda^2 \int_0^{\infty} e^{-x(\lambda - t)} dx \\ &= \frac{\lambda^2}{(\lambda - t)^2} \int_0^{\infty} ue^{-u} du \quad (u = x(\lambda - t)) \quad (t < \lambda) \end{aligned}$$

If $t \geq \lambda$ then the integrals above diverge so the MGF does not exist for these values of t. If $t < \lambda$, the integrals evaluate to 1. So

$$M_z(t) = \frac{\lambda^2}{(\lambda - t)^2} = (1 - \frac{t}{\lambda})^{-2}$$

(b)

$$\begin{aligned} M_{X+Y}(t) &= M_X(t)M_Y(t) \\ &= \frac{\lambda}{\lambda - t} * \frac{\lambda}{\lambda - t} \quad (t < \lambda) \\ &= M_z(t) \end{aligned}$$

The MGF's of X and Y also do not exist when $t \geq \lambda$. So by equality of MGFs we have $Z \sim X + Y$

(10) Let C be a Cauchy random variable, i.e. C has density function

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}.$$

  (a) Show that the moment generating function $M_C(t)$ of C is infinite, except at $t = 0$.
  (b) For which numbers $\alpha > 0$ is $\mathbb{E}[C^\alpha] < \infty$?

(c) Let Z_1, Z_2 be independent r.v.'s with $\mathrm{N}(0,1)$ distribution. Show that Z_1/Z_2 has the same distribution as C.

(d) Let C_1, C_2 be independent r.v.'s with Cauchy distribution. Show that $\frac{1}{2}(C_1 + C_2)$ has the same distribution as C. [Challenging]

Solution: (a) $M_C(t) = \mathbb{E}[e^{tC}] = \int_{-\infty}^{\infty} \frac{1}{\pi} \frac{e^{tx}}{1+x^2}\, dx = \mathrm{sgn}(t) \cdot \infty$ if $t \neq 0$, since

$$\lim_{x\to\infty} \frac{e^{tx}}{1+x^2} = \mathrm{sgn}(t) \cdot \infty$$

,

where $\mathrm{sgn}(\cdot) \in \{\pm 1\}$ is the sign of $\cdot$. Since f is a probability density, $M_C(0) = 1$.

(b) By the p-test, $\mathbb{E}[C^\alpha] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x^\alpha}{1+x^2}\, dx$ is infinite if and only if $\alpha \geq 1$.

(c) Let $W = Z_1/Z_2$, and use the quotient 'convolution' formula to find the PDF of W (similar derivation to the product convolution formula):

$$f_W(w) = \int_{-\infty}^{\infty} |z| f_{Z_1}(wz) f_{Z_2}(z)\, dz = \frac{1}{2\pi} \int_{-\infty}^{\infty} |w| \exp(-\frac{1}{2} w^2 (1+z^2))\, dz = \frac{1}{\pi} \frac{1}{1+z^2}.$$

(d) For a direct computation, see this stackexchange post. Alternatively, one can use a tool similar to MGF's, called 'characteristic functions', or the 'Fourier transform:'

$$\phi_X(t) = M_X(it) = \mathbb{E}[e^{itX}],$$

where $i = \sqrt{-1}$. The characteristic function is an extension of the MGF to the complex numbers, and it shares many of its properties, in particular that if X and Y are independent, then $\phi_{X+Y} = \phi_X \phi_Y$. Importantly, it is well defined for the Cauchy distribution, and one can check that $\phi_{C/2}^2 = \phi_C$, proving the desired property.

(11) Let $\theta_1, \theta_2, \theta_3$ be independent uniform random variables on $[0, 2\pi]$. Let T be the random triangle with vertices on the unit circle at angles $\theta_1, \theta_2, \theta_3$, and let X be the area of T.

(a) Let $\alpha = \min\{\theta_2, \theta_3\}, \beta = \max\{\theta_2, \theta_3\}, \gamma = \beta - \alpha = |\theta_3 - \theta_2|$. Find the PDF's of α, β, γ.

(b) Show that X is equal in distribution to

$$X = \frac{1}{2}\left(\sin\alpha - \sin\beta + \sin\gamma\right).$$

(Hint: assume WLOG that $\theta_1 = 0$.)

(c) Use the expression from part (b) to find $\mathbb{E}X$.

Additional exercises: Anderson 4.20, 6.28, 6.48, 6.58, 8.29, 8.64