1 Problems - Set 1

1.1 Kolmogorov axioms

Show that the definition of *frequentist* probability given in the lectures satisfies the three Kolmogorov axioms (I don't expect any very rigorous proofs). If you can also show this for the Bayesian definition, then have a go (but you don't need to in order to get the full marks). For this question, stick to thinking about probabilities of *exclusive events*.

ANSWER:

Frequentist probability. The definition of frequentist probability given in the lectures is that if counting a particular outcome X as n from a total number of trials N, the ratio n/N converges to the probability for that outcome to occur - or,

$$\lim_{N\to\infty}\frac{n}{N} = P(X)$$

- For the first axiom, clearly both n and N are positive, so the ratio n/N must also be positive for any n. It follows that $P(X) \geq 0$ for any X.
- The second axiom is also satisfied since if one counts either X or Y occurances in N trials as n_X and n_Y, the total number of outcomes either X or Y must be $n_X + n_Y$. If the limits exist, then from the definition of limits, we know,
$$P(X \text{ or } Y) = \lim_{N\to\infty}\frac{n_X+n_Y}{N} = \lim_{N\to\infty}\frac{n_X}{N} + \lim_{N\to\infty}\frac{n_Y}{N} = P(X) + P(Y)$$
- Finally, the third axiom must be satisfied since if one counts any possible outcome X_i as n_{X_i} from N trials, the probability of any outcome to occur must be
$$P(\text{any } X_i) = \lim_{\to\infty}\frac{\sum_i n_{X_i}}{N}.$$
But since $\sum_i n_{X_i} = N$, $P(\text{any } X_i) = \lim_{\to\infty}\frac{N}{N} = 1$.

Bayesian probability. For Bayesian probability, we can use the concept of betting odds. Its easiest to show that the Bayesian definition of probability satisfies the axioms by using a simple game, rolling a six-sided die. Remember we define the probability of a win by,

$$P(W) = \frac{B}{B+W},$$

where B is the amount we bet and W is the *minimum* amount of *profit* we would be happy to win for a bet of B - i.e $B+W$ is the amount we need to *win* such that in the long run (or across different games) we'd expect to break even.

- For the first axiom, clearly the amount we bet or win should be positive so the ratio $B/(B+W)$ must be positive for any B. It follows that $P(W) \geq 0$.
- For the second axiom, let's say we define two games based on rolling a six-sided die. In the first game, we win if the die rolls a 1. In this case, we would need to win £5 for a bet of £1 when we roll a 1. The probability is then $P(1) = \frac{1}{1+5} = \frac{1}{6}$. In a second game, if the die rolls greater or equal to 5, we win and in this case, we would expect $W =$£2 for $B =$£1 and $P(\geq 5) = \frac{1}{1+2} = \frac{1}{3}$. If we combine the games such that we win on either a roll of 1 or a roll of greater than 5, we would expect to $W =$£1 for $B =$£1 to break even (have a think about why). Note that we cannot roll both 1 and greater than 5 in the same game so these are exclusive outcomes. If we sum the probabilities from the individual games, we get, $\frac{1}{6} + \frac{1}{3} = \frac{1}{2} = \frac{1}{1+1} = P(1 \text{ or } \geq 5)$.
- For the third axiom, we are asking how much we would expect to profit if we win on any outcome of the die roll. Clearly, we could ask for 0 profit and still break even in the long run. We have,
$$P(\text{anything}) = \frac{1}{1+0} = 1.$$

1.2 No correlation does not mean independence

In the lectures, we said that two random variables which are independent will have a zero correlation coefficient.

1. Show that two continuous random variables X and Y, with $(X, Y) \sim f(X, Y)$ which are independent will have a 0 correlation coefficient.
2. Let X be a continuous random variable symmetrically distributed around 0 with a density function $f(X)$. Let $Y = X^2$. Show that despite the fact that Y and X are clearly dependent, their correlation coefficient is 0.

ANSWER:

Part 1 The first thing to remember is that if X and Y are independent then their probability distributions can be decomposed into $f(X, Y) = g(X)h(Y)$. The expectation value of (X, Y) is,

$$E[XY] = \int\int XY \cdot f(X,Y)dXdY = \int\int XY \cdot h(X)g(Y)dXdY = \int Xh(X)dX \int Yg(Y)dY = E[X]E[Y].$$

From the lectures, we know that the definition of the correlation coefficient is,

$$\rho(X_i, X_j) = \frac{\nu_{1,1}^{ij}}{\sqrt{v_2^i v_2^j}} = \frac{E[(X_i - E[X_i])(X_j - E[X_j])]}{\sqrt{E[(X_i - E[X_i])^2]E[(X_j - E[X_j])^2]}}$$

Looking at the numerator of the right hand side of this expression, if $\nu_{1,1}^{ij} = 0$, then $\rho(X_i, X_j) = 0$. Multiplying out the outer expectation in the numerator, we have,

$$E[(X_i - E[X_i])(X_j - E[X_j])] = E[X_iX_j - X_iE[X_j] - X_jE[X_i] + E[X_i]E[X_j]],$$

and by the linearity property of the expectation $E[\cdot]$ and the fact that $E[X_i]$ and $E[Xj]$ are constant numbers, we have,

$$E[X_iX_j - X_iE[X_j] - X_jE[X_i] + E[X_i]E[X_j]] = E[X_iX_j] - E[X_i]E[X_j] - E[X_j]E[X_i] + E[X_i]E[X_j]E[1] = E[X_iX_j] - E[X_i]E[X_j].$$

We already showed that $E[XY] = E[X]E[Y]$ if X and Y are independent, so we have that the correlation coefficient is zero.

Part 2 Again we can calculate just the numerator ($\nu_{1,1}^{ij}$) of the correlation coefficient formula. We have,

$$\nu_{1,1}^{X,X^2} = E[X \cdot X^2] - E[X]E[X^2].$$

Since X is *symmetrically distributed around 0*, the expectation of X is 0, and so is any odd function (any function such that $g(-X) = -g(X)$) of X such as $f(X) = X^3$. We can see this since the expectation value of the function $g(X)$ is given by,

$$\begin{aligned}
\int_{-\infty}^{\infty} g(X)f(X)dX &= \int_{-\infty}^{0} g(X)f(X)dX + \int_{0}^{\infty} g(X)f(X)dX \\
&= \int_{-\infty}^{0} g(X)f(-X)dX + \int_{0}^{\infty} g(X)f(X)dX \\
&= -\int_{-\infty}^{0} g(-X)f(-X)dX + \int_{0}^{\infty} g(X)f(X)dX \\
&= \int_{-\infty}^{0} g(-X)f(-X)(-dX) + \int_{0}^{\infty} g(X)f(X)dX \\
&= \int_{\infty}^{0} g(X)f(X)dX + \int_{0}^{\infty} g(X)f(X)dX \\
&= -\int_{0}^{\infty} g(X)f(X)dX + \int_{0}^{\infty} g(X)f(X)dX \\
&= 0
\end{aligned}$$

where we've used the fact that $f(X) = f(-X)$ i.e it is the (symmetric or even) probability distribution of X. Therefore, $\nu_{1,1}^{X,X^2}$ and hence the correlation between X and X^2 is zero. Clearly these two are very much dependent so this example shows how zero correlation does not necessarily imply independence between two variables.

1.3 Moments of the Poisson distribution

In the lecture notes, we calculated the expectation and variance of the Poisson distribution with parameter λ are $E[k] = \lambda$ and $V(k) = \lambda$.

What is the third central moment of the Poisson distribution? You should show how you derive your answer.

ANSWER: In the lecture notes, we had an example where we calculated the expectation and variance of the Poisson distribution. We used the following identities there;

$$\sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} = (\lambda^2 + \lambda)e^{\lambda}$$

and

$$\sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} = \lambda e^{\lambda}$$

The definition of the *third central moment* of a discrete random variable k with probability distribution $p(k)$ was defined in lectures to be,

$$E\left[(k - E[k])^3\right] = \sum_k (k - E[k])^3 p(k).$$

For the case where $p(k)$ is a Poisson distribution with parameter λ this becomes,

$$E\left[(k - E[k])^3\right] = \sum_{k=0}^{\infty}(k-\lambda)^3\lambda^k\frac{e^{-\lambda}}{k!},$$

where we used the fact that $E[k] = \lambda$ for the Poisson case, as shown in the lectures.

Now, lets expand the brackets,

$$\begin{aligned}
E\left[(k - E[k])^3\right] = & \textstyle\sum_{k=0}^{\infty}(k-\lambda)^3\lambda^k\frac{e^{-\lambda}}{k!} \\
= & \textstyle\sum_{k=0}^{\infty} k^3\lambda^k\frac{e^{-\lambda}}{k!} \\
+ & 3\lambda^2\textstyle\sum_{k=0}^{\infty} k\lambda^k\frac{e^{-\lambda}}{k!} \\
- & 3\lambda\textstyle\sum_{k=0}^{\infty} k^2\lambda^k\frac{e^{-\lambda}}{k!} \\
- & \lambda^3\textstyle\sum_{k=0}^{\infty} \lambda^k\frac{e^{-\lambda}}{k!}.
\end{aligned}$$

The last term is easy since this is just $\lambda^2\sum_k p(k) = \lambda^2$ as the total probability must equal 1. For the 2nd and 3rd terms, we can use the identities above to simplify them so that we are left with,

$$\begin{aligned}
E\left[(k - E[k])^3\right] = & \textstyle\sum_{k=0}^{\infty}(k-\lambda)^3\lambda^k\frac{e^{-\lambda}}{k!} \\
= & \textstyle\sum_{k=0}^{\infty} k^3\lambda^k\frac{e^{-\lambda}}{k!} \\
+ & 3\lambda^2 \times \lambda \\
- & 3\lambda \times (\lambda^2 + \lambda) \\
- & \lambda^3.
\end{aligned}$$

For the first term then, we need to think a bit more. We can notice however that,

$$\begin{aligned}
\sum_{k=0}^{\infty} k^3\lambda^k\frac{e^{-\lambda}}{k!} = & \; 0 + \textstyle\sum_{k=1}^{\infty} k^3\lambda^k\frac{e^{-\lambda}}{k!} \\
= & \textstyle\sum_{k=1}^{\infty}(\cancel{k} \times k^2)(\lambda \times \lambda^{k-1})\frac{e^{-\lambda}}{\cancel{k} \times (k-1)!} \\
= & \; \lambda\textstyle\sum_{k=1}^{\infty} k^2\lambda^{k-1}\frac{e^{-\lambda}}{(k-1)!} \\
= & \; \lambda\textstyle\sum_{p=0}^{\infty}(p+1)^2\lambda^p\frac{e^{-\lambda}}{p!},
\end{aligned}$$

where we've made the substitution $k \to p + 1$ in the last line. Now we can expand the brackets to get,

$$\lambda\sum_{p=0}^{\infty}(p+1)^2\lambda^p\frac{e^{-\lambda}}{p!} = \lambda\left(\sum_{p=0}^{\infty}p^2\lambda^p\frac{e^{-\lambda}}{p!} + \sum_{p=0}^{\infty}2p\lambda^p\frac{e^{-\lambda}}{p!} + \sum_{p=0}^{\infty}\lambda^p\frac{e^{-\lambda}}{p!}\right),$$

and once again we can simplify using our identities,

$$\lambda\sum_{p=0}^{\infty}(p+1)^2\lambda^p\frac{e^{-\lambda}}{p!} = \lambda\left(\lambda^2 + \lambda + 2\lambda + 1\right) = \lambda^3 + 3\lambda^2 + \lambda,$$

and finally we can substitute this back into our expression for the third moment to get to,

$$\begin{aligned}
E\left[(k - E[k])^3\right] = & \textstyle\sum_{k=0}^{\infty}(k-\lambda)^3\lambda^k\frac{e^{-\lambda}}{k!} \\
= & \; \lambda^3 + 3\lambda^2 + \lambda + 3\lambda^3 - 3\lambda^3 - 3\lambda^2 - \lambda^3 \\
= & \; \lambda,
\end{aligned}$$

so the third central moment of a Poisson distribution is λ.

Alternative Answer: Another way of finding the answer is to use moment generating functions and cumulant generating functions. We didn't discuss these at all in the lectures but if you already know about them, it's fine to use this method for this question.

The moment generating function for a Poisson random variable is given by,

$$m(t) = e^{\lambda(e^t - 1)} \tag{1}$$

and the cumulant generating function is given by the logarithm of the moment generating function, so for a Poisson it is,

$$c(t) = \lambda(e^t - 1). \tag{2}$$

It is generally true that the first algebraic moment and, second and third central moments are equal to the first, second and third derivatives of the cumulative generating function evaluated at $t = 0$, so we have that,

$$\nu_3 = \frac{d^3 c(t)}{dt^3}|_{t=0} = \lambda e^0 = \lambda. \tag{3}$$

1.4 Cauchy distribution

In the lectures, we showed how the sum of two Gaussian distributed random variables is itself Gaussian distributed. Suppose now that $X \sim \phi(X; 0, 1)$ and $Y \sim \phi(Y; 0, 1)$ are independent random variables. Show that the distribution of $U = \dfrac{X}{Y}$ is Cauchy, i.e that $p(U) = \dfrac{1}{\pi(1 + U^2)}$.

Hint: You should start by deriving the marginal distribution formula for the ratio of two independent random variables. Careful that the case $Y = 0$ will cause a problem, so split the marginal distribution into two cases, one for $Y > 0$ and one for $Y < 0$. The sum of these marginal distributions will be the total distribution.

ANSWER: As the hint in the question suggested, we start by deriving the marginal distribution for the ratio of our two independent variables. Let $U = \frac{X}{Y}$ and $V = Y$ as our change of variables. For $Y \neq 0$, our infitesimal volume transforms as,

$$dXdY = \left|\det\begin{bmatrix} \frac{\partial U}{\partial X} & \frac{\partial U}{\partial Y} \\ \frac{\partial V}{\partial X} & \frac{\partial V}{\partial Y} \end{bmatrix}\right|^{-1} dUdV = \left|\det\begin{bmatrix} \frac{1}{Y} & -\frac{X}{Y^2} \\ 0 & 1 \end{bmatrix}\right|^{-1} dUdV = Y \cdot dUdV = V \cdot dUdV.$$

We calculate the marginal distribution $p(U)$ first for values of V greater than zero. We have,

$$\begin{aligned} p(U)^+ = \int_\epsilon^\infty h(U, V)dV &= \int_\epsilon^\infty \phi(UV)\phi(V)VdV \\ &= \int_\epsilon^\infty \frac{1}{2\pi} e^{-\frac{U^2V^2}{2}} e^{-\frac{V^2}{2}} VdV \\ &= \int_\epsilon^\infty \frac{1}{2\pi} e^{-\frac{V^2(1+U^2)}{2}} VdV \\ &= -\frac{1}{2\pi}\left[\frac{1}{1+U^2} e^{-\frac{V^2(1+U^2)}{2}}\right]_\epsilon^\infty \\ &= \frac{1}{2\pi(1+U^2)} e^{-\frac{\epsilon^2(1+U^2)}{2}} \end{aligned}$$

where $\epsilon > 0$ is some negligible number and the superscript $^+$ means we only take the part of the probability distribution from the region where $Y = V > 0$. In the limit of $\epsilon \to 0$, we get,

$$p(U)^+ = \frac{1}{2\pi(1+U^2)}.$$

Similarly, we can take the case where $Y = V < 0$ (for $p(U)^-$ and we will get an identical result so that,

$$p(U) = p(U)^- + p(U)^+ = \frac{1}{\pi(1+U^2)},$$

which is the Cauchy distribution.