ECOM40006/90013 Econometrics 3

Week 1 Extras

Question 1: Expectations and Variance

(a) Let $\mu := \mathbb{E}(X)$ to reduce notational burden. Then expanding out the variance formula gives

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}((X-\mu)^2) \\
&= \mathbb{E}(X^2 - 2X\mu + \mu^2) \\
&= \mathbb{E}(X^2) - 2\mathbb{E}(X\mu) + \mathbb{E}(\mu^2) \\
&= \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 && \because\ \mu \text{ is constant} \\
&= \mathbb{E}(X^2) - 2\mu^2 + \mu^2 \\
&= \mathbb{E}(X^2) - \mu^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2,
\end{aligned}$$

as required.

(b) Now define $\mu_X := \mathbb{E}(X)$ and $\mu_Y := \mathbb{E}(Y)$. The covariance formula, when expanded, gives

$$\begin{aligned}
\mathrm{cov}(X,Y) &= \mathbb{E}((X-\mu_X)(Y-\mu_Y)) \\
&= \mathbb{E}(XY - X\mu_Y - \mu_X Y + \mu_X\mu_Y) \\
&= \mathbb{E}(XY) - \mu_Y\mathbb{E}(X) - \mu_X\mathbb{E}(Y) + \mu_X\mu_Y && \because\ \mu_X, \mu_Y \text{ constant} \\
&= \mathbb{E}(XY) - 2\mu_X\mu_Y + \mu_X\mu_Y \\
&= \mathbb{E}(XY) - \mu_X\mu_Y = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)
\end{aligned}$$

as required.

(c) The answer is no, since the general formula for the variance of the sum of two RVs is

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{cov}(X,Y).$$

From inspection, we can only claim that $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ whenever $\mathrm{cov}(X,Y) = 0$. That is: X and Y are uncorrelated RVs.

(d) Straight calculation gives us

$$\begin{aligned}
\mathrm{Var}(AZ_1) &= \mathbb{E}((AZ_1 - \mathbb{E}(AZ_1))(AZ_1 - \mathbb{E}(AZ_1))') \\
&= \mathbb{E}((AZ_1 - A\mathbb{E}(Z_1))(AZ_1 - A\mathbb{E}(Z_1))') \\
&= \mathbb{E}(A(Z_1 - \mathbb{E}(Z_1))[A(Z_1 - \mathbb{E}(Z_1))]') \\
&= \mathbb{E}(A(Z_1 - \mathbb{E}(Z_1))(Z_1 - \mathbb{E}(Z_1))'A') && \because (AB)' = B'A' \\
&= A\,\underbrace{\mathbb{E}((Z_1 - \mathbb{E}(Z_1))(Z_1 - \mathbb{E}(Z_1))')}_{\mathrm{Var}(Z_1)}\,A' \\
&= A\mathrm{Var}(Z_1)A',
\end{aligned}$$

as required. Keep in mind that $\mathbb{E}(X)' = \mathbb{E}(X')$ as well when doing these calculations on your own.

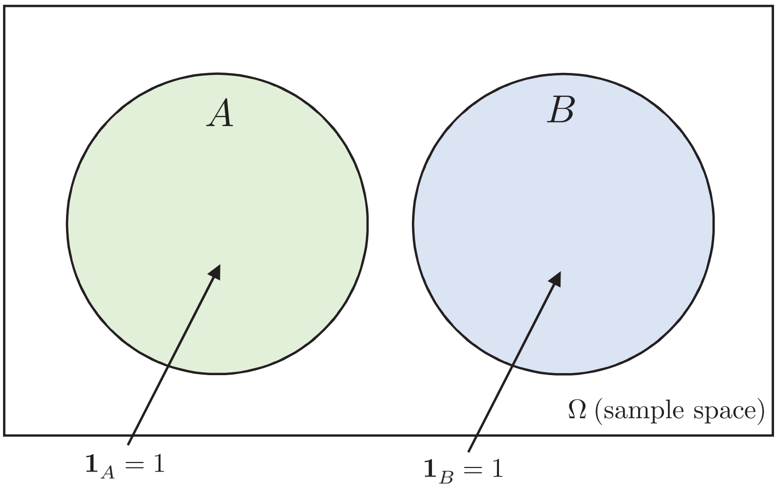
(e) Provided that we have the covariance formula for two random vectors on hand, expansion in a very similar fashion to part (d) above gives us

$$\begin{aligned}
\mathrm{cov}(AZ_1, BZ_2) &= \mathbb{E}\left((AZ_1 - A\mathbb{E}(Z_1))(BZ_2 - B\mathbb{E}(Z_2))'\right) \\
&= \mathbb{E}(A[Z_1 - \mathbb{E}(Z_1)][Z_2'B' - \mathbb{E}(Z_2)'B']) \\
&= A\,\underbrace{\mathbb{E}([Z_1 - \mathbb{E}(Z_1)][Z_2 - \mathbb{E}(Z_2)]')}_{\mathrm{cov}(Z_1,Z_2)=\Sigma}\,B' \\
&= A\Sigma B',
\end{aligned}$$

as required. Note that in the first line, we pre-emptively made use of the fact that $\mathbb{E}(AZ_1) = A\mathbb{E}(Z_1)$ and similarly $\mathbb{E}(BZ_2) = B\mathbb{E}(Z_2)$ since both A and B are non-stochastic.

Question 2: Indicator Random Variables

(a) Below is the classic Venn diagram that you might be familiar with from the past. Anything that falls into the shaded area marked A, for example, is classified as 'event A occurs.' In indicator variable notation, we translate 'A occurs' into $\mathbf{1}_A = 1$. Similarly, we have that $\mathbf{1}_B = 1$ whenever B occurs. This is always true regardless of whether or not the events are disjoint.



(b) The key point to convincing yourself that these statements are true is to keep in mind that having an indicator variable equal 1 is synonymous to having that event occurring. A quick argument for each of these cases might also supplement the Venn diagrams themselves.

(i.) Whenever $\mathbf{1}_A = 1$, calculating $\mathbf{1}_{A^c} = 0$. In other words, when A happens, its complement A^c does not occur. In the other case, when $\mathbf{1}_A = 0$, then the expression $\mathbf{1}_{A^c} = 1$, indicating that when A does not occur, then its complement A^c occurs.

Hence, this expression is only equal to 1 whenever A^c occurs and 0 when it does not, which is consistent with what we have been doing.

(ii.) Following the same argument as before, let's consider a few cases.

- If A occurs and B does not, we have $\mathbf{1}_A\mathbf{1}_B = 1 \times 0 = 0$.
- Similarly, $\mathbf{1}_A\mathbf{1}_B = 0$ if B occurs and A doesn't.
- If both A and B occur, then $\mathbf{1}_A\mathbf{1}_B = 1 \times 1 = 1$.
- If neither A nor B occur, then we have $0 \times 0 = 0$.

So this expression behaves exactly as we would expect from an indicator variable for $\mathbf{1}_{A\cap B}$.

(iii.) Let's try the same argument as before:

- If A occurs and B doesn't, then we have

$$\max\{\mathbf{1}_A, \mathbf{1}_B\} = \max\{1, 0\} = 1.$$

- Similarly, we have $\max\{0, 1\} = 1$ for when B occurs and A doesn't.
- If neither A nor B occurs, then we have $\max\{0, 0\} = 0$.
- If both A and B occur, we have $\max\{1, 1\} = 1$.

All in all, this also behaves the same way that we expect from the indicator for the event $A \cup B$.

(c) First, note that

$$\begin{aligned}\mathbb{E}(\mathbf{1}_A) &= 1 \times \Pr(\mathbf{1}_A) + 0 \times (1 - \Pr(\mathbf{1}_A)) \\ &= \Pr(\mathbf{1}_A).\end{aligned}$$

Namely: the expected value of an indicator random variable is the probability with which $\mathbf{1}_A$ equals 1. Alternatively, you can think of the expected value of $\mathbf{1}_A$ as simply being the probability with which A occurs. In other words:

$$\Pr(\mathbf{1}_A) = \Pr(\mathbf{1}_A = 1) = \Pr(A),$$

since $\mathbf{1}_A = 1$ only when A occurs. So note that

$$\begin{aligned}\operatorname{cov}(\mathbf{1}_A, \mathbf{1}_B) &= \mathbb{E}(\mathbf{1}_A\mathbf{1}_B) - \mathbb{E}(\mathbf{1}_A)\mathbb{E}(\mathbf{1}_B) \\ &= \mathbb{E}(\mathbf{1}_{A\cap B}) - \Pr(A)\Pr(B) \\ &= \Pr(A \cap B) - \Pr(A)\Pr(B).\end{aligned}$$

Now, if A and B are independent, then $\Pr(A\cap B) = \Pr(A)\Pr(B)$ and hence $\operatorname{cov}(\mathbf{1}_A, \mathbf{1}_B) = 0$.

Since this is a bidirectional proof, we should also go the other way. Note that if we begin with the premise that $\operatorname{cov}(\mathbf{1}_A, \mathbf{1}_B) = 0$ then our results above immediately imply that

$$\Pr(A \cap B) - \Pr(A)\Pr(B) = 0 \implies \Pr(A \cap B) = \Pr(A)\Pr(B).$$

This is the definition of independent RVs, so if the indicator RVs have zero covariance, then A and B are independent events.

(d) We know the following about the indicator RV:

- $\mathbf{1}_A = 1$ occurs with probability p
- $\mathbf{1}_A = 0$ occurs with probability $1 - p$.

The goal is to get this into a nicer functional form. As it turns out, we can use the probability mass function for what is called a *Bernoulli* random variable:

$$\Pr(\mathbf{1}_A = x) = p^x(1-p)^{1-x}, \qquad x = 0, 1.$$

Notice that if $x = 0$, then this function is equal to $1 - p$. If $x = 1$, then we have the function being equal to p instead.

(e) We already found out earlier that the mean of an indicator RV is simply the probability with which A occurs:

$$\mathbb{E}(\mathbf{1}_A) = \Pr(A).$$

All we need to do is find the variance, then. The formula for the variance of a discrete RV gives us

$$\begin{aligned}
\operatorname{Var}(\mathbf{1}_A) &= p(1-p)^2 + (1-p)(0-p)^2 \\
&= p(1-p)^2 + p^2(1-p) \\
&= (1-p)[p(1-p) + p^2] \\
&= (1-p)[p - p^2 + p^2] \\
&= p(1-p).
\end{aligned}$$

(f) We have $S_2 = X_1 + X_2$. We already figured out the mean and variance for a single X_i, so we may use this to our advantage. One then has

$$\begin{aligned}
\mathbb{E}(S_2) &= \mathbb{E}(X_1) + \mathbb{E}(X_2) \\
&= 2\mathbb{E}(X_1) && \text{(Identical distributions)} \\
&= 2p
\end{aligned}$$

For the variance, we have

$$\begin{aligned}
\operatorname{Var}(S_2) &= \operatorname{Var}(X_1 + X_2) \\
&= \operatorname{Var}(X_1) + \operatorname{Var}(X_2) && \text{(Independence)} \\
&= 2\operatorname{Var}(X_1) && \text{(Identical distributions)} \\
&= 2p(1-p).
\end{aligned}$$

If we generalize this to the sum

$$S_n = \sum_{i=1}^{n} X_i$$

then we find that

$$\mathbb{E}(S_n) = np, \qquad \operatorname{Var}(S_n) = np(1-p).$$

Fun fact: these are the mean and variance respectively for a binomial random variable!

Question 3: Moment Generating Functions

(a) Beginning from $f(X) = e^{tX}$, notice that if you take repeated derivatives, you will find that the n^{th} derivative can be written as

$$f^{(n)}(X) = t^n e^{tX} \qquad \text{so} \qquad f^{(n)}(0) = t^n.$$

Therefore,[1]

$$\begin{aligned}
f(X) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!}(X-0)^n \\
&= \sum_{n=0}^{\infty} \frac{t^n}{n!} X^n.
\end{aligned}$$

[1] A conceptual abuse occurs here in the sense that the Taylor series of a function is not always equal to the function itself. We gloss over this fact for the sake of focusing on the underlying results, but we make note of it here for those who wish for a little bit more rigor.

(b) Note that

$$\mathbb{E}(f(X)) = \mathbb{E}\left(\sum_{n=0}^{\infty} \frac{t^n}{n!} X^n\right) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}(X^n) = 1 + t\mathbb{E}(X) + \frac{t^2}{2}\mathbb{E}(X^2) + \dots$$

The Taylor series expansion of this function tells us that if we knew every instance of $\mathbb{E}(X^n)$, then we could calculate $\mathbb{E}(e^{tX})$.

(c) If we denote $M_X(t) := \mathbb{E}(e^{tX})$ then further observe that

$$M_X'(t) = \frac{dM_X(t)}{dt} = \mathbb{E}(X) + t\mathbb{E}(X^2) + \frac{t^2}{2}\mathbb{E}(X^3) + \dots$$

and that $M_X'(0) = \mathbb{E}(X)$. It turns out that the n^{th} derivative of this function, evaluated at $t = 0$, will give back the n^{th} raw moment $\mathbb{E}(X^n)$. Such functions are called *moment generating functions* and can be very useful for deriving particular properties of random variables.

(d) Note that for this question, taking the mean and variance of Y directly will yield the desired answers, but doing so doesn't answer the question of whether or not Y is also *normal*. For this we'll need to use the MGF of the normal distribution that is provided.

First, let's look at the MGF of Y. By definition,

$$M_Y(t) = \mathbb{E}[e^{tY}].$$

But since $Y = aX + b$ we can just substitute that in:

$$\begin{aligned} M_Y(t) &= \mathbb{E}[e^{t(aX+b)}] \\ &= \mathbb{E}[e^{atX+bt}] \\ &= \mathbb{E}[e^{atX}e^{bt}] \\ &= e^{bt}\mathbb{E}[e^{(at)X}] \end{aligned}$$

since e^{bt} is constant. Note that at is also a number so $\mathbb{E}[e^{(at)X}]$ is also a moment generating function of X, just evaluated at at instead of t:

$$\begin{aligned} M_Y(t) &= e^{bt}M_X(at) \\ &= e^{bt}e^{\mu(at)+\frac{\sigma^2(at)^2}{2}} \\ &= e^{a\mu t+bt+\frac{a^2\sigma^2t^2}{2}} \\ &= e^{(a\mu+b)t+\frac{(a^2\sigma^2)t^2}{2}} \\ &= e^{\tilde{\mu}t+\frac{\tilde{\sigma}^2t^2}{2}}, \end{aligned}$$

where $\tilde{\mu} = a\mu + b$ and $\tilde{\sigma}^2 = a^2\sigma^2$. Since the form of the MGF matches that for a $N(\tilde{\mu}, \tilde{\sigma}^2)$ distribution we can conclude that $Y = aX + b$ is normally distributed with mean $a\mu + b$ and variance $a^2\sigma^2$. In other words,

$$Y \sim N(a\mu + b, a^2\sigma^2).$$