MATH7501 exam 2013 — solutions and marking scheme

1. (a) PGF is $\Pi(z) = E(z^X)$.
Probability generating functions are defined for discrete random variables taking non-negative integer values.

From the definition of expectation, we have $\Pi(z) = \sum_{k=0}^{\infty} z^k p(k)$ where $p(\cdot)$ is the probability mass function of X. Thus

$$\begin{aligned}\Pi'(z) &= \frac{d}{dz}\left[p(0) + zp(1) + z^2p(2) + \dots\right] \\ &= p(1) + 2zp(2) + 3z^2p(3) + \dots\end{aligned}$$

So $\Pi'(1) = p(1) + 2p(2) + 3p(3) + \dots = \sum_{k=0}^{\infty} kp(k) = E(X)$, as required.

Marking note: *A simpler solution would be to write $\Pi'(z) = E(Xz^{X-1})$. This works, but makes the implicit assumption that differentiation and expectation can be interchanged. Full marks will be given for any solution that goes down this route, even though we haven't formally shown in the course that such an interchange is legitimate.*
Next, we have $\Pi''(z) = 2p(2) + 3.2zp(3) + 4.3z^2p(4) + \dots$, so that

$$\Pi''(1) = 2p(2) + 3.2p(3) + 4.3p(4) + \dots = \sum_{k=0}^{\infty} k(k-1)p(k) = E(X(X-1)) ,$$

as required (note that the first two terms in the final sum here are zero).

Marking note: *alternative solution uses $\Pi''(z) = E(X(X-1)z^{X-2})$ — again, full marks for this.*

(b) The probability mass function of X is obtained from the coefficients in the series expansion of $\Pi(z)$. Here we have

$$\begin{aligned}\Pi(z) = \left(\frac{1}{4}\right)^n (3z+1)^n &= \left(\frac{1}{4}\right)^n \sum_{k=0}^{n} \binom{n}{k} (3z)^k \\ &= \sum_{k=0}^{n} \binom{n}{k} (3/4)^k (1/4)^{n-k} z^k .\end{aligned}$$

Reading off the coefficients in this expansion therefore, the pmf of X is

$$P(X=k) = \begin{cases} \binom{n}{k}(3/4)^k(1/4)^{n-k} & k = 0, 1, \dots, n \\ 0 & \text{otherwise.} \end{cases}$$

For the expectation, we have $\Pi'(z) = 3n(1/4)^n(3z+1)^{n-1}$ so that $\Pi'(1) = 3n/4 = E(X)$.

For the variance, we have $\Pi''(z) = 3^2n(n-1)(1/4)^n(3z+1)^{n-2}$ so that $\Pi''(1) = (3/4)^2n(n-1) = E[X(X-1)]$. Therefore, writing $\mu = E(X)$, we have

$$\begin{aligned}\operatorname{Var}(X) = E[X(X-1)] - \mu(\mu-1) &= (3/4)^2n(n-1) - (3n/4)(3n/4-1) \\ &= \frac{3n}{4}\left[\frac{3(n-1)}{4} - \frac{3n-4}{4}\right] = \frac{3n}{16}.\end{aligned}$$

The distribution of X is Binomial(n,3/4).

---

2. (a) Suppose first that $g(\cdot)$ is strictly increasing. Then the distribution function of Y is

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) = P\left(X \le g^{-1}(y)\right) = F_X\left(g^{-1}(y)\right) .$$

Differentiating, the density of Y is

$$f_Y(y) = \frac{d}{dy}\left[F_X\left(g^{-1}(y)\right)\right] = f_X\left(g^{-1}(y)\right)\frac{d}{dy}\left(g^{-1}(y)\right) = f_X\left(g^{-1}(y)\right)\frac{dx}{dy} .$$

If $g(\cdot)$ is strictly increasing then $dx/dy > 0$ so that $|dx/dy| = dx/dy$. Therefore the result holds when $g(\cdot)$ is strictly increasing.

If $g(\cdot)$ is instead strictly *decreasing*, we have

$$P(g(X) \le y) = P\left(X \ge g^{-1}(y)\right) = 1 - F_X\left(g^{-1}(y)\right) ,$$

and in this case the density of Y is $-f_X\left(g^{-1}(y)\right)dx/dy$. But if $g(\cdot)$ is strictly decreasing then dx/dy is negative, so $-dx/dy = |dx/dy|$ and the density of Y can be written as $f_X\left(g^{-1}(y)\right)|dx/dy|$. Thus the result also holds when $g(\cdot)$ is strictly decreasing.

Marking note: *5 marks for the strictly increasing case, 3 marks for dealing adequately with the strictly decreasing case.*

The need for $g(\cdot)$ to be strictly monotonic arises from the step $P(g(X) \le y) = P\left(X \begin{smallmatrix}\le\\ \ge\end{smallmatrix} g^{-1}(y)\right)$ — this requires that the inverse function $g^{-1}(\cdot)$ exists and is uniquely defined.

(b) The density of U is $f_U(u) = 1$ for $u \in (0,1)$ and zero otherwise. Thus, if $Y = g(U)$ then the density of Y is just $|du/dy|$ over the values taken by Y.

i. Put $Y = -3\log(1-U) = g(U)$, where $g(u) = -3\log(1-u)$ so that $g^{-1}(y) = 1 - e^{-y/3}$. $g(\cdot)$ is monotonic increasing; and Y takes values in

$\mathbb{R}^+$. Moreover, $du/dy = d/dy\,(g^{-1}(y)) = \frac{1}{3}e^{-y/3}$. Thus the density of Y is

$$f_Y(y) = \begin{cases} \frac{1}{3}e^{-y/3} & y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

This is the density of the exponential distribution with parameter $\lambda = 1/3$.

Marking note*: 4 marks for the derivation, 1 for naming the distribution (with parameter). For full marks in the derivation, need to consider explicitly the range of values taken by Y. Similarly for the other two examples.*

ii. Put $Y = \Phi^{-1}(U) - 7 = g(U)$, where $g(u) = \Phi^{-1}(u) - 7$ so that $g^{-1}(y) = \Phi(y+7)$. $g(\cdot)$ is monotonic increasing; and Y takes values in $\mathbb{R}$. Moreover, $du/dy = d/dy\,(g^{-1}(y)) = \phi(y+7)$ where $\phi(\cdot)$ is the pdf of the standard normal distribution. Thus the density of Y is

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(y+7)^2}{2}\right] \qquad (y \in \mathbb{R})\ .$$

This is the density of the normal distribution with parameters $\mu = -7$ and $\sigma^2 = 1$.

iii. Put $Y = U^2 = g(U)$, where $g(u) = u^2$ so that $g^{-1}(y) = y^{1/2}$. $g(\cdot)$ is monotonic increasing over the range of U; and Y takes values in $(0,1)$. Moreover, $du/dy = d/dy\,(g^{-1}(y)) = y^{-1/2}/2$. Thus the density of Y is

$$f_Y(y) = \begin{cases} \frac{1}{2}y^{-1/2} & y \in (0,1) \\ 0 & \text{otherwise.} \end{cases}$$

This is the density of the beta distribution with parameters $\alpha = 1/2$ and $\beta = 1$.

---

3. (a) The MGF of Y is $M_Y(t) = E(e^{tY})$

$$= E\left[e^{t(aX+b)}\right] = E\left(e^{tb}e^{atX}\right) = e^{tb}E\left(e^{atX}\right) = e^{tb}M_X(at)\ ,$$

as required.

(b) The density of $N(0,1)$ is $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, for $x \in \mathbb{R}$.[1] Thus, if $X \sim N(0,1)$ then the MGF is

$$\begin{aligned} M(t) = E(e^{tX}) &= \int_{\mathbb{R}} e^{tx}\phi(x)dx = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left[tx - \frac{x^2}{2}\right] dx \\ &= \text{(completing the square)} \end{aligned}$$

[1]This formula is in the statistical tables provided, for those students who have forgotten it — although the ones who forget are probably the least likely to realise this.

$$
\begin{aligned}
&\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left[\frac{t^2}{2} - \frac{t^2}{2} + tx - \frac{x^2}{2}\right] dx \\
=\;& \exp\left[\frac{t^2}{2}\right] \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x-t)^2\right] dx \;.
\end{aligned}
$$

The integrand here is the density of $N(t,1)$ and hence the integral is 1; thus we have $M(t) = e^{t^2/2}$, as required.

The density of $\Gamma(\alpha, \lambda)$ is $f(x) = \frac{1}{\Gamma(\alpha)} x^{\alpha-1}\lambda^\alpha e^{-\lambda x}$, for $x > 0$. Thus, if $X \sim \Gamma(\alpha, \lambda)$, the MGF is

$$
\begin{aligned}
M(t) = E(e^{tX}) &= \int_0^\infty e^{tx} f(x) dx = \int_0^\infty \frac{1}{\Gamma(\alpha)} x^{\alpha-1}\lambda^\alpha e^{-(\lambda-t)x} dx \\
&= \frac{\lambda^\alpha}{(\lambda-t)^\alpha} \int_0^\infty \frac{1}{\Gamma(\alpha)} x^{\alpha-1}(\lambda-t)^\alpha e^{-(\lambda-t)x} dx \;.
\end{aligned}
$$

But, providing $t < \lambda$ so that $\lambda - t > 0$, the integrand here is the density of $\Gamma(\alpha, \lambda - t)$ and hence the integral is 1. Therefore, for $t < \lambda$ we have $M(t) = \lambda^\alpha(\lambda - t)^{-\alpha}$ as required.

Marking note: *4 marks here are for the basic derivation; the remaining mark is for dealing adequately with the need for $t < \lambda$.*

(c) Using the results from part (a) with $a = \lambda$ and $b = -\alpha$, the MGF of $\lambda X_i - \alpha$ is $e^{-t\alpha}\lambda^\alpha(\lambda - \lambda t)^{-\alpha} = e^{-t\alpha}(1-t)^{-\alpha}$. Next, using the fact that the MGF of a sum of independent random variables is the product of their MGFs, the MGF of $\sum_{i=1}^n (\lambda X_i - \alpha)$ is $e^{-nt\alpha}(1-t)^{-n\alpha}$. Finally, using the results from part (a) again with $a = (n\alpha)^{-1/2}$ and $b = 0$, the MGF of Z_n is $M_n(t) = e^{-(n\alpha)^{1/2}t}\left[1 - (n\alpha)^{-1/2}t\right]^{-n\alpha}$.

We therefore have

$$
\begin{aligned}
\log M_n(t) &= -(n\alpha)^{1/2}t - n\alpha \log\left[1 - (n\alpha)^{-1/2}t\right] \\
&= -(n\alpha)^{1/2}t + n\alpha \sum_{k=1}^\infty \frac{1}{k}\left[(n\alpha)^{-1/2}t\right]^k \;,
\end{aligned}
$$

providing $|(n\alpha)^{-1/2}t| < 1$, which is OK since we're considering $n \to \infty$. Writing the sum out longhand, we get

$$
\begin{aligned}
\log M_n(t) &= -(n\alpha)^{1/2}t + n\alpha\left[(n\alpha)^{-1/2}t + \frac{1}{2}(n\alpha)^{-1}t^2 + \frac{1}{3}(n\alpha)^{-3/2}t^3 + \ldots\right] \\
&= -(n\alpha)^{1/2}t + (n\alpha)^{1/2}t + \frac{1}{2}t^2 + \frac{1}{3}(n\alpha)^{-1/2}t^3 + \ldots \;.
\end{aligned}
$$

This series obviously converges by the ratio test, for n large enough. Thus, as $n \to \infty$ we have $\log M_n(t) \to t^2/2$, and hence $M_n(t) \to e^{t^2}/2$. This is the MGF of $N(0,1)$; since the MGF uniquely characterises the distribution, the distribution of Z_n tends to $N(0,1)$ as required.

Marking note: *testing for convergence not required to obtain full marks.*

---

4. (a) i. $p(x,y) = P(X = x, Y = y) = P(X = x)P(Y = y|X = x)$. The joint pmf is therefore

$p(x,y)$		y	
		0	1
x	0	$(1-p)(1-q_0)$	$(1-p)q_0$
	1	$p(1-q_1)$	pq_1

ii. Y takes values 0 and 1, so its marginal distribution is Bernoulli. We have $P(Y = 1) = (1-p)q_0 + pq_1$. So $Y \sim Ber\left((1-p)\,q_0 + pq_1\right)$.

(b) Let X and Y be two random variables. The Iterated Expectation Law states that

$$E(Y) = E_X\left[E_{Y|X}(Y|X)\right],$$

where $E_X(\cdot)$ denotes expectation over the marginal distribution of X and $E_{Y|X}(\cdot|\cdot)$ denotes expectation over the conditional distribution of Y given the individual values taken by X.

Marking note: *5 marks may seem generous here, however there is plenty of scope for inaccuracy / imprecision, and giving 5 marks in total provides the opportunity to award, say, 3 out of 5 for an answer that is a bit confused over notation but nonetheless conveys the general sense. Obviously, full marks will be awarded if X and Y are interchanged.*

In the example from part (a), $E(Y) = (1-p)\,q_0 + pq_1$ (left-hand side of iterated expectation law). We also have $E_{Y|X}(Y|X = 0) = q_0$, and $E_{Y|X}(Y|X = 1) = q_1$. Since $P(X = 1) = p = 1 - P(X = 0)$, the right-hand side of the law thus reads

$$E_X\left[E_{Y|X}(Y|X)\right] = q_0(1-p) + q_1 p,$$

which is exactly the required result.

(c) i. If day 0 is dry then $X_1 \sim Ber(1/4)$ (marginal distribution of X_1)

Putting $p = 1/4$, $q_0 = 1/4$ and $q_1 = 2/3$ in the result from part (a), we now find that the joint distribution of X_1 and X_2 is

		x_2	
		0	1
x_1	0	9/16	3/16
	1	1/12	1/6

So the marginal distribution of X_2 is $Ber(3/16+1/6) = Ber(17/48)$ (or $Ber(0.354)$).

ii. To find the marginal distribution of X_3, we can repeat exactly the same procedure starting with the joint distribution of X_2 and X_3. Alternatively, as a short cut we can note that X_3 is Bernoulli and that

$$\begin{aligned} P(X_3=1) &= P(X_3=1|X_2=0)P(X_2=0)+P(X_3=1|X_2=1)P(X_2=1) \\ &= (1/4\times 31/48)+(2/3\times 17/48) = 229/576 \text{ (or } 0.398) \ . \end{aligned}$$

So $X_3 \sim Ber(229/576)$ (or $\text{Ber}(0.398)$).

iii. If $X_1 \sim Ber(p)$ then $X_2 \sim Ber((1-p)/4+2p/3) = Ber(\frac{3+5p}{12})$, using the result from part (a)(ii). We thus require

$$p = \frac{3+5p}{12} \Rightarrow 12p = 3+5p \Rightarrow p = \frac{3}{7} = 0.429 \ .$$

---

5. (a) i. Let $\mu = E(T)$. Then the MSE of T is defined as

$$\begin{aligned} \text{MSE}(T) = E\left[(T-\theta)^2\right] &= E\left[(T-\mu+\mu-\theta)^2\right] \\ &= E\left[(T-\mu)^2+(\mu-\theta)^2+2\,(T-\mu)\,(\mu-\theta)\right] \\ &= E\left[(T-\mu)^2\right]+(\mu-\theta)^2+2E\left[(T-\mu)\,(\mu-\theta)\right] \ . \end{aligned}$$

Now $E\left[(T-\mu)^2\right] = \text{Var}(T)$; $\mu-\theta = b(T)$; and

$$E\left[(T-\mu)\,(\mu-\theta)\right] = (\mu-\theta)\,E\,(T-\mu) = 0 \ .$$

Hence

$$\text{MSE}(T) = \text{Var}(T) + b^2(T) \ ,$$

as required.

ii. T_n is consistent for θ if, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\left(|T_n-\theta| > \varepsilon\right) = 0 \ .$$

Marking note: *2 marks out of 3 for saying that T_n is consistent if $\lim_{n\to\infty} MSE(T_n) = 0$ — this is sufficient but not necessary.*

(b) i. $E(\overline{X}_n) = \mu$; $E(\overline{Y}_n) = 2\mu$; $\text{Var}(\overline{X}_n) = \sigma^2/n$; and $\text{Var}(\overline{Y}_n) = 2\sigma^2/n$.

ii. For T_n to be unbiased for μ then we need $E(T_n) = \mu$ i.e.

$$E\left(a\overline{X}_n + b\overline{Y}_n\right) = \mu \qquad \Rightarrow \qquad aE\left(\overline{X}_n\right) + bE\left(\overline{Y}_n\right) = a\mu + 2b\mu = \mu\ ,$$

so that $a + 2b = 1$ and $a = 1 - 2b$, as required.
The variance of T_n is

$$\text{Var}(T_n) = a^2\text{Var}(\overline{X}_n) + b^2\text{Var}(\overline{Y}_n) = \frac{(1-2b)^2\sigma^2}{n} + \frac{2b^2\sigma^2}{n}\ .$$

For minimum variance we therefore need to minimise $(1 - 2b)^2 + 2b^2 = 6b^2 - 4b + 1$, which has a minimum at $b = 1/3$.

iii. It is of interest to minimise the variance of T_n because it is unbiased and in this case minimising the variance is equivalent to minimising the mean squared estimation error by the result in part (a)(i). Thus, by minimising the variance we obtain in some sense the most precise unbiased estimator possible.

iv. Putting $b = 1/3$ and $a = 1 - 2/3$ in the expression for $\text{Var}(T_n)$ above, we find

$$\text{Var}(T_n) = \frac{\sigma^2}{n}\left[\left(\frac{1}{3}\right)^2 + 2\left(\frac{1}{3}\right)^2\right] = \frac{\sigma^2}{3n}\ .$$

By contrast, $\text{Var}(\overline{X}_{2n}) = \sigma^2/2n$. Since $\sigma^2/3n < \sigma^2/2n$, it is better to use the estimator from part (ii).

---

6. (a) If the data are paired then the two groups of observations cannot be considered as independent. The derivation of the two-sample test involves the variance of the difference between the two sample means, $\overline{X}$ and $\overline{Y}$ say. If the two groups are independent then this variance is $\text{Var}(\overline{X})+\ \text{Var}(\overline{Y})$; and this is used to standardise the difference between the two means when constructing the t-statistic. If the two groups are not independent however, the variance of the difference is no longer the sum of the individual variances and hence the standardisation is incorrect.

Marking note: *just 2 marks for saying "the two groups are not independent", because this doesn't follow through to explain how this affects the derivation.*

In practice, in a paired setting it is likely that the two observations within a pair will be more similar to each other than to observations for other pairs. As a result, differences between the two groups may be smaller than the differences between the pairs; and if this is not taken into account properly, the effect of interest may be swamped by inter-pair variation. The consequence is that the paired test may fail to detect a genuine difference between the two groups.

Marking note: *any other sensible / aware comments will be rewarded. Fatuous waffle will be ignored (including answers that say "the results will be misleading" without explaining what this means). Arrant nonsense will be penalised.*

σ_b^2

(b) i. Under $H_0 : \cancel{\sigma_n^2} = \sigma_g^2$, the test statistic $F = s_b^2/s_g^2$ follows an F distribution with 24 and 15 degrees of freedom. We will therefore reject H_0 if F falls in either the upper or lower 2.5% of this distribution. Since $s_b^2 > s_g^2$ here, we only need to check the upper 2.5% point, which is 2.701 from Table 12(c) of the supplied tables.
The observed value of F is $8^2/7^2 = 1.3061 < 2.701$. Hence we accept H_0 and conclude that the data are consistent with the hypothesis that $\sigma_b^2 = \sigma_g^2$, as required.

ii. Since the variances can be assumed equal, we can perform a 2-sample t-test. The test statistic is

$$T = \frac{\bar{x}_b - \bar{x}_g}{s_P\sqrt{25^{-1} + 16^{-1}}} ,$$

where s_P^2 is the pooled sample variance:

$$s_P^2 = \frac{24s_b^2 + 15s_g^2}{39} .$$

Under $H_0 : \mu_b = \mu_g$ (where μ_b and μ_g denote the underlying means for boys and girls respectively), T follows a t-distribution with 39 degrees of freedom. We will therefore reject H_0 if $|T|$ exceeds the upper 2.5% of this distribution, and accept otherwise.
Table 10 of the supplied tables gives the upper 2.5% points of t_{38} and t_{40} as 2.024 and 2.021 respectively. The upper 2.5% point of t_{39} is therefore approximately 2.0225.
For the data given here, we have $s_P^2 = [(24 \times 8^2) + (15 \times 7^2)]/39 = 58.231$. We therefore have

$$T = \frac{72 - 68}{\sqrt{58.231(25^{-1} + 16^{-1})}} = \frac{4}{\sqrt{5.969}} = 1.637 .$$

Since $1.637 < 2.0225$, we accept H_0 and conclude that these data do not provide any evidence for a difference between the mean maths exam scores of boys and girls.