MATH7501 Examination 2012: Solutions and Marking Scheme

1. (a) Axioms are

   $P(E) \geq 0$,

   $P(\Omega) = 1$,

   if $E \cap F = \emptyset$ then $P(E \cup F) = P(E) + P(F)$.

   (b) i. A random variable is a function from the sample space to the real numbers: $X : \Omega \to \mathbb{R}$.

   ii. Let $F(x) = P(X \leq x)$ be the distribution function of a random variable X. X is continuous iff $F(x)$ is continuous and differentiable $\forall x$.

   (c) $P(X \leq b) = P((X \leq a) \cup (a < X \leq b))$ when $b > a$. This is equal to $P(X \leq a) + P(a < X \leq b)$ (axiom 3).

   But $P(X \leq b) = F(b)$ and $P(X \leq a) = F(a)$ by definition.

   Hence $P(a < X \leq b) = F(b) - F(a)$ as required.

   From above, we have that $F(b) = F(a) + P(a < X \leq b)$ $(b > a)$ and $P(a < X \leq b) \geq 0$ (axiom 1). Hence, when $b > a$ we have $F(b) \geq F(a)$ and $F(\cdot)$ is nondecreasing as required.

   (d) Let $E_1, \ldots, E_n$ be a partition of Ω, i.e. a collection of mutually disjoint subsets with $\cup_{j=1}^n E_j = \Omega$. Let F be any event. Bayes' Theorem states that for each i

   $$P(E_i|F) = \frac{P(F|E_i)P(E_i)}{\sum_{j=1}^n P(F|E_j)P(E_j)}.$$

   Proof:

   $P(E_i|F) = P(E_i \cap F)/P(F)$ by definition of conditional probability.

   Similarly, $P(F|E_i) = P(F \cap E_i)/P(E_i)$ so that $P(F \cap E_i) = P(F|E_i)P(E_i)$ and hence $P(E_i|F) = P(F|E_i)P(E_i)/P(F)$.

   Now, $P(F) = P(F \cap \Omega) = P(F \cap (\cup_{j=1}^n E_j))$ which is equal to $P(\cup_{j=1}^n (F \cap E_j))$ by distributive laws and then equal to $\sum_j P(F \cap E_j)$ by countable additivity. Finally this is equal to $\sum_j P(F|E_j)P(E_j)$.

2. (a) For f to be the probability density function of a continuous random variable, we require that:

$f(x) \geq 0 \ \forall x$,

$\int_{\mathbb{R}} f(x)dx = 1$.

(b) F is the distribution function of a continuous random variable if:

$F(-\infty) = 0$,

$F(\infty) = 1$,

F is non-decreasing,

F is everywhere continuous.

(c) $F(x) = \int_{-\infty}^{x} f(u)du = ab \int_0^x u^{b-1} \exp(-au^b)du \ (x > 0)$.

If $v = au^b$ and $dv = bau^{b-1}du$ then:

$$\int_0^{ax^b} e^{-v}dv = [-e^{-v}]_0^{ax^b} = 1 - e^{-ax^b}(x > 0).$$

Note that $F(x) = 0$ for $x \leq 0$.

The distribution function of Y is

$P(Y \leq y) = P(X^b \leq y) = P(X \leq y^{1/b}) = F(y^{1/b}) = 1 - e^{-ay}$.

This is the distribution function of an exponential distribution with parameter a.

3. (a) The density is

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & otherwise \end{cases}.$$

The corresponding distribution function is

$$F(x) = \int_{-\infty}^{x} f(u)du = \begin{cases} 0 & x < 0 \\ x & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}.$$

(b) The distribution function of X is

$$F_X(x) = P(X \le x) = P(-\log U^2 \le x) = P(\log U \ge -\frac{x}{2}) = P(U \ge e^{-\frac{x}{2}})$$
$$1 - F(e^{-\frac{x}{2}}) = 1 - e^{-\frac{x}{2}} \quad (x > 0).$$

This is the cumulative distribution function of an exponential distribution with parameter 1/2.

The density of X is

$$f(x) = \frac{1}{2}e^{-\frac{x}{2}} \quad (x > 0).$$

For $t \le 1/2$, the MGF is

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx}\frac{1}{2}e^{-\frac{x}{2}}dx = \frac{1}{2}\left[\frac{e^{(t-\frac{1}{2})x}}{t-\frac{1}{2}}\right]_0^\infty = (1-2t)^{-1}$$

(c) $S = -\sum_{i=1}^n \log(U_i^2) = \sum_{i=1}^n X_i$, where $X_i \sim Exp(\frac{1}{2})$ independently for each i. Hence $M_S(t) = \prod_{i=1}^n M_X(t) = (1-2t)^{-n}$, for $t \le 1/2$.

(d) Comparing $M_S(t)$ with the expression given, we have equality if $n = m/2$. Hence $m = 2n$ and $S \sim \chi^2_{2n}$.

(e) $P(\prod_{i=1}^6 U_i > 0.1) = P(\prod_{i=1}^6 U_i^2 > 0.01) = P(-\sum_{i=1}^6 \log U_i^2 < -\log 0.01) = P(Y < 4.61)$, where $Y \sim \chi^2_{2\times 6}$. This is equal to $(0.8\times 0.0274)+(0.2\times 0.0420) = 0.030$ by linear interpolation in the corresponding table.

4. (a)

$$E\left[e^{aX}\right] = \sum_{k=0}^{\infty} e^{ak} P(X=k) = \sum_{k=0}^{\infty} e^{ak} \frac{m^k e^{-m}}{k!} = e^{-m} \sum_{k=0}^{\infty} \frac{(me^a)^k}{k!} = e^{-m} e^{me^a} = e^{m(e^a-1)}.$$

(b) i. Since $S_n \sim Poi(n\mu)$, we have that

$$E(e^{-n^{-1}S_n}) = e^{n\mu(e^{-n^{-1}}-1)}.$$

This is obtained from the part (a) with $m = n\mu$ and $a = -n^{-1}$. It follows that

$$E(T) = 1 - e^{n\mu(e^{-n^{-1}}-1)} \neq p.$$

So T is biased for p.

Since

$$n\mu(e^{-n^{-1}} - 1) = n\mu\left(-n^{-1} + \frac{n^{-2}}{2!} - \dots\right) = -\mu + \frac{\mu}{2n} - \dots,$$

as $n \to \infty$ we must have $n\mu(e^{-n^{-1}} - 1) \to -\mu$. Hence,

$$E(T) \to 1 - e^{-\mu} = p.$$

ii. $Y \sim Bin(n,p)$. $E(Y) = np$ and $Var(Y) = np(1-p)$. Hence $E(Y/n) = p$ which is unbiased for p.
The standard error of this estimator is $\sqrt{Var(Y/n)} = \sqrt{p(1-p)/n}$. **(unseen)** [5]

iii. A good estimator should have a small bias and variance. Now, T is biased and Y/n is not. T may be preferable if its bias is small and its variance is less than that of Y/n. To choose between these two estimators, we would therefore need to calculate the variance of T. A way to combine bias and variance of an estimator is the mean squared error: $MSE = bias^2 + variance$. The estimator with the smallest MSE would be preferred in general.

5. (a) i. $\overline{X} \sim N(\mu_1, \sigma_1^2/m)$.

ii. $\overline{X} - \overline{Y} \sim N(\mu_1 - \mu_2, \sigma_1^2/m + \sigma_2^2/n)$.

iii.

$$Z = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}}.$$

(b) i. Under $H_0 : \sigma_1^2 = \sigma_2^2$, the test statistic $F = S_1^2/S_2^2$ is distributed as $F_{10,8}$. From the tables, the upper 2.5% point of this is 4.295. The lower 2.5% point is

$$\frac{1}{upper\ 2.5\%\ point\ of\ F_{8,10}} = \frac{1}{3.855} = 0.259.$$

The observed value of F is $0.039/0.023 = 1.69$. Since this is between 0.259 and 4.295, we do not reject H_0 and may conclude that the data are consistent with the two groups having equal variance.

ii. Under $H_0 : \mu_1 = \mu_2$, the test statistic

$$t = \frac{\overline{X} - \overline{Y}}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

is distributed as $t_{n_1+n_2-2}$, where $n_1 = 11$ and $n_2 = 9$. The upper and lower 0.5% points of t_{18} are ± 2.878 (from the tables). Hence we will not reject H_0 if $|t| < 2.878$.

The observed value of t is

$$t = \frac{0.71 - 1.06}{S_p\sqrt{\frac{1}{11} + \frac{1}{9}}} = -\frac{0.35}{S_p 0.499},$$

where $S_p = \sqrt{\frac{10S_1^2 + 8S_2^2}{18}} = 0.179$. Hence the observed value of t is $-0.35/(0.499 \times 0.179) = -3.918$. Since $|3.918| > 2.878$ we reject H_0 and conclude that at the 1% level there is evidence for a difference in the underlying means of the two samples.

iii. The F test suggests that the training course has not affected the variability of the performance. However, the t test suggests that it has affected the mean. Although a 2-sided test was used, it seems reasonable to conclude that the training course has improved the average workers' performance.

6. (a) $E(\overline{X}) = aE(X_1) + bE(X_2) = a\mu + b\mu$, where μ is the true weight (since measurement errors have mean zero). We need $(a+b)\mu = \mu$ so that $b = 1-a$ for $\mu \neq 0$ as required.

(b) Since the errors are independent, $Var(\overline{X}) = a^2Var(X_1)+b^2Var(X_2) = a^2\sigma_1^2 + (1-a)^2\sigma_2^2$. For a minimum, differentiate wrt a and set to zero:

$$2a\sigma_1^2 - 2(1-a)\sigma_2^2 = 0 \quad \Rightarrow \quad a = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}.$$

This must be a minimum since $Var(\overline{X})$ is quadratic in a with a positive leading term.

If $\overline{X}$ has small variance then it is likely to be close to its expected value, hence the measurement error is likely to be small.

(c) i. When $\sigma_1^2 = 4$ and $\sigma_2^2 = 1$, we set $a = 1/5$ and $b = 4/5$ so that $Var(\overline{X}) = 4/25 + 16/25 = 4/5$. Now, $\overline{X}$ is a normal since X_1 and X_2 both are. Hence, $\overline{X} \sim N(50, 4/5)$.

ii.

$$P(\overline{X} < 48) = P(\frac{\overline{X}-50}{\sqrt{4/5}} < \frac{48-50}{\sqrt{4/5}}) = P(Z < -2.236) = 0.0127,$$

where $Z \sim N(0,1)$.]

iii. The probability sought is $P(49 < \overline{X} < 51)$. This is equal to $1 - 2P(\overline{X} > 51) = 1 - 2P(Z > 1.118) = 0.737$.]