Queen’s University
Department of Mathematics and Statistics

STAT 353
Final Examination April 9, 2009
Instructor: G. Takahara

• “Proctors are unable to respond to queries about the interpretation of exam questions. Do your best to answer exam questions as written.”

• “The candidate is urged to submit with the answer paper a clear statement of any assumptions made if doubt exists as to the interpretation of any question that requires a written answer.”

• Formulas and tables are attached.

• An 8.5 × 11 inch sheet of notes (both sides) is permitted.

• Simple calculators are permitted. HOWEVER, do reasonable simplifications.

• Write the answers in the space provided, continue on the backs of pages if needed.

• SHOW YOUR WORK CLEARLY. Correct answers without clear work showing how you got there will not receive full marks.

• Marks per part question are shown in brackets at the right margin.

Marks: Please do not write in the space below.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Marks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem 1</td>
<td>10</td>
</tr>
<tr>
<td>Problem 2</td>
<td>10</td>
</tr>
<tr>
<td>Problem 3</td>
<td>10</td>
</tr>
<tr>
<td>Problem 4</td>
<td>10</td>
</tr>
<tr>
<td>Problem 5</td>
<td>10</td>
</tr>
<tr>
<td>Problem 6</td>
<td>10</td>
</tr>
</tbody>
</table>

Total: 60
1. Let $X$ and $Y$ be independent random variables, where $X$ has a Poisson distribution with parameter 1 and $Y$ has an exponential distribution with parameter 1. Show that

$$E \left[ \left( \frac{Y}{2} \right)^X \right] = \frac{2}{e}. $$

Hint: You can condition on $X$ and compute $E[Y^n]$ for any nonnegative integer $n$. [10]

Solution: Conditioning on $X = n$, we have

$$E \left[ \left( \frac{Y}{2} \right)^X \bigg| X = n \right] = E \left[ \left( \frac{Y}{2} \right)^n \bigg| X = n \right] = E \left[ \left( \frac{Y}{2} \right)^n \right] = \frac{1}{2^n} E[Y^n],$$

since $X$ and $Y$ are independent. Computing $E[Y^n]$, we have

$$E[Y^n] = \int_0^\infty y^n e^{-y} dy = n! \int_0^\infty \frac{1}{n!} y^n e^{-y} dy = n!$$

where the last equality follows because the integrand is the pdf of a gamma distribution with parameters $n + 1$ and 1. Therefore, by the law of total expectation

$$E \left[ \left( \frac{Y}{2} \right)^X \right] = \sum_{n=0}^{\infty} \frac{n!}{2^n} P(X = n) = \sum_{n=0}^{\infty} \frac{n!}{2^n} \frac{1}{n!} e^{-1} = e^{-1} \sum_{n=0}^{\infty} \frac{1}{2^n} = \frac{2}{e},$$

as desired.
2. (a) Let $X$ and $Y$ be independent random variables each having the uniform distribution on $(0, 1)$. Let $U = \min(X, Y)$ and $V = \max(X, Y)$. Compute $\text{Cov}(U, V)$. 

**Hint:** Note that $UV = XY$ with probability 1. [6]

**Solution:** First, from the hint

$$E[UV] = E[XY] = E[X]E[Y] = \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) = \frac{1}{4},$$

where the second equality follows since $X$ and $Y$ are independent. From the formula sheet the pdf of $U = \min(X, Y)$ is

$$f_U(u) = \begin{cases} 2(1-u) & \text{for } 0 < u < 1 \\ 0 & \text{otherwise} \end{cases}$$

and the pdf of $V = \max(X, Y)$ is

$$f_V(v) = \begin{cases} 2v & \text{for } 0 < v < 1 \\ 0 & \text{otherwise}. \end{cases}$$

Therefore,

$$E[U] = \int_0^1 2u(1-u)du = 2 \left[ \frac{u^2}{2} - \frac{u^3}{3} \right]_0^1 = \frac{2}{6} = \frac{1}{3},$$

and

$$E[V] = \int_0^1 2v^2dv = 2 \left[ \frac{v^3}{3} \right]_0^1 = \frac{2}{3}.$$

Thus, we get

$$\text{Cov}(U, V) = E[UV] - E[U]E[V] = \frac{1}{4} - \left(\frac{1}{3}\right) \left(\frac{2}{3}\right) = \frac{1}{4} - \frac{2}{9} = \frac{1}{36}.$$
(b) Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with finite variance, and let $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. Show that $\text{Cov}(\overline{X}_n, X_i - \overline{X}_n) = 0$ for all $i = 1, \ldots, n$. [4]

Solution: Using properties of covariance, we have (for $i$ fixed)

$$\text{Cov}(\overline{X}_n, X_i - \overline{X}_n) = \text{Cov} \left( \frac{1}{n} \sum_{j=1}^{n} X_j, \frac{1}{n} \sum_{k=1}^{n} (X_i - X_k) \right)$$

$$= \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \text{Cov}(X_j, X_i - X_k)$$

$$= \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \text{Cov}(X_j, X_i) - \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} \text{Cov}(X_j, X_k)$$

$$= \frac{1}{n} \text{Cov}(X_i, X_i) - \frac{1}{n^2} \sum_{j=1}^{n} \text{Cov}(X_j, X_j)$$

$$= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0,$$

where $\sigma^2 = \text{Var}(X_i)$, many of the terms in going from the third line above to the fourth line are equal to zero because $X_1, \ldots, X_n$ are independent, and $\text{Cov}(X_j, X_j) = \text{Var}(X_j) = \sigma^2$. 
3. Let $Y$ be a continuous random variable with probability density function $f_Y(y)$ and moment generating function $M_Y(t)$ (assume $|M_Y(t)| < \infty$ for all $t$). For a fixed $\tau$, let $X$ have probability density function

\[ f_X(x) = \frac{e^{\tau x} f_Y(x)}{M_Y(\tau)}. \]

(a) Find $M_X(t)$, the moment generating function of $X$. [4]

**Solution:** The moment generating function of $X$ is

\[ M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx = \frac{1}{M_Y(\tau)} \int_{-\infty}^{\infty} e^{(t+\tau)x} f_Y(x) dx = \frac{M_Y(t + \tau)}{M_Y(\tau)}. \]

(b) Suppose that $Y \sim N(\mu, \sigma^2)$. Find $E[X]$. [6]

**Solution:** If $Y \sim N(\mu, \sigma^2)$, then the moment generating function of $Y$ is

\[ M_Y(t) = e^{t\mu + \sigma^2 t^2/2} \]

and so the moment generating function of $X$ (from part (a)) is

\[ M_X(t) = \frac{M_Y(t + \tau)}{M_Y(\tau)} = \frac{e^{(t+\tau)\mu + \sigma^2(t+\tau)^2/2}}{e^{t\mu + \sigma^2 t^2/2}} = e^{t\mu + \sigma^2 t^2/2 + \sigma^2 t \tau}. \]

The mean of $X$ can then be computed as

\[ E[X] = M_X'(0) = \left[(\mu + \sigma^2 t + \sigma^2 \tau)e^{t\mu + \sigma^2 t^2/2 + \sigma^2 t \tau}\right]_{t=0} = \mu + \sigma^2 \tau. \]
4. Consider a sequence of independent experiments, where in each experiment we take \( k \) balls, labelled 1 to \( k \) and randomly place them into \( k \) slots, also labelled 1 to \( k \), so that there is exactly one ball in each slot. For the \( i \)th experiment, let \( X_i \) be the number of balls whose label matches the slot label of the slot into which it is placed. So \( X_1, X_2, \ldots \) is a sequence of independent and identically distributed random variables.

   (a) Find the mean and variance of \( X_i \). *Hint:* Write \( X_i = X_{i1} + \ldots + X_{ik} \), where \( X_{ij} \) is the indicator that slot \( j \) receives ball \( j \) in the \( i \)th experiment.

   **Solution:** Writing \( X_i = X_{i1} + \ldots + X_{ik} \) as suggested in the hint, we have that

   \[
   E[X_i] = \sum_{j=1}^{k} P(X_{ij} = 1),
   \]

   where \( P(X_{ij} = 1) \) is the probability that the slot with label \( j \) receives the ball with label \( j \). Since all assignments of the balls to the slots are equally likely, a simple counting argument gives that

   \[
   P(X_{ij} = 1) = \frac{(k-1)!}{k!} = \frac{1}{k}.
   \]

   Therefore, \( E[X_i] = \sum_{j=1}^{k} \frac{1}{k} = 1 \). Similarly,

   \[
   E[X_i^2] = E[(X_{i1} + \ldots + X_{ik})^2] = \sum_{j=1}^{k} E[X_{ij}^2] + \sum_{j \neq r} E[X_{ij}X_{ir}] = 1 + \sum_{j \neq r} E[X_{ij}X_{ir}],
   \]

   where the last equality follows since \( X_{ij}^2 = X_{ij} \) (for indicator random variables). Another counting argument yields

   \[
   E[X_{ij}X_{ir}] = P(X_{ij} = 1, X_{ir} = 1) = P(\text{ball } j \text{ goes in slot } j \text{ and ball } r \text{ goes in slot } r) = \frac{(k-2)!}{k!} = \frac{1}{k(k-1)}.
   \]

   Since there are \( k(k-1) \) terms in the final sum in Eq.(1), we have \( E[X_i^2] = 2 \), and so

   \[
   \text{Var}(X_i) = E[X_i^2] - E[X_i]^2 = 2 - 1^2 = 2 - 1 = 1.
   \]
(b) Use the central limit theorem to approximate the probability that in the first 25 experiments the total number of balls whose label matches their slot label is greater than 30. [4]

Solution: The total number of balls in the first 25 experiments whose label matches their slot label is \( \sum_{i=1}^{25} X_i \), where \( X_1 \ldots, X_{25} \) are independent and identically distributed random variables and where (from part (a)) we have that \( E[X_i] = 1 \) and \( \text{Var}(X_i) = 1 \). By the central limit theorem

\[
P\left( \sum_{i=1}^{25} X_i > 30 \right) = P\left( \frac{\sum_{i=1}^{25} X_i - 25}{5} > \frac{30 - 25}{5} \right) \approx P(Z > 1),
\]

where \( Z \sim N(0, 1) \). That is,

\[
P\left( \sum_{i=1}^{25} X_i > 30 \right) \approx 1 - \Phi(1) = 1 - 0.8413 = 0.1587.
\]
5. Let $Y_1, Y_2, \ldots$ be a sequence of discrete random variables such that the joint probability mass function of $(Y_1, \ldots, Y_n)$ is

$$P(Y_1 = y_1, \ldots, Y_n = y_n) = \begin{cases} 
(n + 1)\left(\frac{1}{n}\right)^{n-1} & \text{for } y_i \in \{0, 1\}, i = 1, \ldots, n \\
0 & \text{otherwise}
\end{cases}$$

(so the $Y_i$’s are Bernoulli random variables but they are not independent). Let $X_n$ be the sample mean of $Y_1, \ldots, Y_n$, i.e., $X_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$. Show that $X_n$ converges in distribution to a limit $X$ and find the distribution of $X$. Hint: $X$ is not a constant. Hint: First find the distribution of $\sum_{i=1}^{n} Y_i$, which has sample space $\{0, 1, \ldots, n\}$. [10]

Solution: Let $S_{k,n} = \{(y_1, \ldots, y_n) : y_i \in \{0, 1\} \text{ and } \sum_{i=1}^{n} y_i = k\}$. Then $\sum_{i=1}^{n} Y_i$ is constant on $S_{k,n}$ (equal to $k$), $S_{k,n}$ contains $\binom{n}{k}$ elements, and

$$P\left(\sum_{i=1}^{n} Y_i = k\right) = P((Y_1, \ldots, Y_n) \in S_{k,n})$$

$$= \sum_{(y_1, \ldots, y_n) \in S_{k,n}} \left[ (n + 1)\left(\frac{1}{n}\right)^{n-1} \right]^{n-1}$$

$$= \sum_{(y_1, \ldots, y_n) \in S_{k,n}} \left[ (n + 1)\left(\binom{n}{k}\right)^{-1} \right]$$

$$= \binom{n}{k} \left(\frac{n + 1}{n}\right)^{n-1} = \frac{1}{n+1}.$$ 

In other words, $\sum_{i=1}^{n} Y_i$ has a discrete uniform distribution on $\{0, 1, \ldots, n\}$. Now let $x \in [0, 1]$. Then

$$P(X_n \leq x) = P\left(\sum_{i=1}^{n} Y_i \leq nx\right) = \frac{\lfloor nx \rfloor + 1}{n + 1} = \frac{nx - c(n, x) + 1}{n + 1} \to x$$

as $n \to \infty$, where $c(n, x)$ is some value satisfying $0 \leq c(n, x) < 1$ for all $n$ and all $x \in [0, 1]$. Clearly, $P(X_n \leq x) = 1$ for all $x > 1$ and $P(X_n \leq x) = 0$ for all $x < 0$. Therefore, the cdf of $X_n$ converges to the cdf of a uniform distribution on $(0, 1)$, at all $x$. In other words, $X_n \to X$ in distribution, where $X \sim U(0, 1)$. 
Formula Sheet

Special Distributions

Continuous uniform on \((a, b)\):

\[
f(x) = \begin{cases} 
1/(b - a) & \text{if } a \leq x \leq b \\
0 & \text{otherwise},
\end{cases}
\]

\[E[X] = \frac{a + b}{2}, \quad \text{Var}[X] = \frac{(b - a)^2}{12}.\]

Exponential with parameter \(\lambda\):

\[
f(x) = \begin{cases} 
\lambda e^{-\lambda x} & \text{if } x > 0 \\
0 & \text{otherwise}.
\end{cases}
\]

\[E[X] = \frac{1}{\lambda}, \quad \text{Var}[X] = \frac{1}{\lambda^2}.\]

Gamma with parameters \(r\) and \(\lambda\):

\[
f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \quad \text{if } x > 0
\]

\[E[X] = \frac{r}{\lambda}, \quad \text{Var}[X] = \frac{r}{\lambda^2}.\]

Normal (Gaussian) with mean \(\mu\) and variance \(\sigma^2\):

\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{and} \quad F(x) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt.
\]

Poisson with parameter \(\lambda\):

\[
P(X = k) = \begin{cases} 
\frac{\lambda^k}{k!} e^{-\lambda} & k = 0, 1, \ldots \\
0 & \text{otherwise}.
\end{cases}
\]

\[E[X] = \lambda, \quad \text{Var}[X] = \lambda.\]

- df and pdf of the kth order statistic from a random sample \(X_1, \ldots, X_n\):

\[
F_k(x) = \sum_{i=k}^{n} \binom{n}{i} [F(x)]^i [1 - F(x)]^{n-i};
\]

\[
f_k(x) = \frac{n!}{(k-1)!(n-k)!} f(x) [F(x)]^{k-1} [1 - F(x)]^{n-k},
\]

where \(F(x)\) and \(f(x)\) are the df and pdf, respectively, of each \(X_i\).
The distribution function of a standard normal random variable