**BIRKBECK**
**(University of London)**

**MSc EXAMINATION FOR INTERNAL STUDENTS**

**MSc STATISTICS etc**

**School of Economics, Mathematics, and Statistics**

# PROBABILITY THEORY

# ANSWERS

**200805011035**

---

1. [**This question is essentially bookwork.**] Let $X_1, X_2, \ldots$ be independent, identically distributed Bernoulli random variables for which $\mathbb{P}(X_i = \pm 1)$, for all $i$, and define the scaled average

$$A_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i.$$

(a) Show that $\mathbb{E}A_n = 0$ and $\mathbb{E}(A_n^2) = 1$.

**2 pts**

**ANS:** Firstly, $\mathbb{E}X_i = (1/2)(1 + (-1)) = 0$, which implies that the scaled sum $A_n$ also satisfies $\mathbb{E}A_n = 0$. Further, $\mathbb{E}(X_i^2) = (1/2)(1+1) = 1$, so the independence of $X_1, X_2, \ldots$ implies

$$\begin{aligned} \mathbb{E}(A_n^2) &= \frac{1}{n}\left(\sum_{i=1}^{n} \mathbb{E}(X_i^2) + \sum_{i \neq j} \mathbb{E}(X_i X_j)\right) \\ &= 1 + \frac{1}{n} \sum_{i \neq j} (\mathbb{E}X_i)(\mathbb{E}X_j) \\ &= 1. \end{aligned}$$

It's also fine if students observe that the variance of a sum of independent random variables is the sum of their variances.

(b) State Chebyshev's inequality and apply it to show that

$$\mathbb{P}(|A_n| \geq t) \leq \frac{1}{t^2},$$

for any $t > 0$.

**3 pts**

**ANS:** Chebyshev's inequality states that, if $V$ is any random variable for which the mean $\mathbb{E}V = \mu$ and variance $\mathbb{E}[(V - \mu)^2] = \sigma^2$ exist, then

$$\mathbb{P}(|V - \mu| \geq t) \leq \frac{\sigma^2}{t^2}, \qquad t > 0.$$

Applying this to $V = A_n$ gives the stated inequality.

(c) Using the Markov inequality

$$e^{ct}\mathbb{P}(A_n \geq t) \leq \mathbb{E}e^{cA_n},$$

for any $c > 0$ and $t \in \mathbb{R}$, prove that

$$\mathbb{P}(A_n \geq t) \leq e^{-ct} \left(\frac{e^{c/\sqrt{n}} + e^{-c/\sqrt{n}}}{2}\right)^n. \qquad (*)$$

**3 pts**

**ANS:** Markov's inequality implies

$$\mathbb{P}(A_n \geq t) \leq e^{-ct}\mathbb{E}e^{cA_n},$$

and, by independence of the $\{X_i\}$,

$$\begin{aligned}
\mathbb{E}e^{cA_n} &= \mathbb{E}\prod_{i=1}^{n} e^{(c/\sqrt{n})X_i} \\
&= \prod_{i=1}^{n} \mathbb{E}e^{(c/\sqrt{n})X_i} \\
&= \prod_{i=1}^{n}\left(\frac{e^{c/\sqrt{n}}+e^{-c/\sqrt{n}}}{2}\right) \qquad = \left(\frac{e^{c/\sqrt{n}}+e^{-c/\sqrt{n}}}{2}\right)^{n}.
\end{aligned}$$

(d) Derive the inequality

$$\frac{e^x+e^{-x}}{2} \le e^{x^2/2}, \qquad \text{for } x \ge 0,$$

using the Taylor series of the exponential function. Hence use (*) to show that

$$\mathbb{P}\left(A_n \ge t\right) \le e^{-ct+c^2/2}. \qquad (**)$$

**4 pts**

**ANS:** We have

$$e^{x^2/2} - (e^x+e^{-x})/2 = \sum_{k=0}^{\infty} x^{2k}\left(\frac{1}{2^k k!} - \frac{1}{(2k)!}\right). \qquad (\dagger)$$

Further,

$$(2k)! = (2k)(2k-1)(2k-2)\cdots 3{\cdot}2{\cdot}1 \ge (2k)(2k-2)(2k-4)\cdots 4{\cdot}2 \ge 2^k k!,$$

which implies that every Taylor coefficient in (†) is non-negative.

(e) Using (**), derive the Bernstein–Azuma–Hoeffding inequality:

$$\mathbb{P}\left(|A_n| \ge t\right) \le 2e^{-t^2/2}$$

**4 pts**

**ANS:** Inequality (**) is valid for any $c > 0$, so we choose $c$ to minimize the upper bound of (**). Now

$$\frac{c^2}{2} - ct = \frac{1}{2}\left((c-t)^2 - t^2\right) \ge -t^2,$$

the minimum occurring when $c = t$, so that

$$\mathbb{P}\left(A_n \ge t\right) \le e^{-t^2/2}.$$

Further,

$$\mathbb{P}\left(A_n \le -t\right) \le e^{-t^2/2},$$

by symmetry of the distribution of $A_n$ (or by minor modification of the above derivation). Hence

$$\mathbb{P}\left(|A_n| \ge t\right) = \mathbb{P}\left(A_n \ge t\right) + \mathbb{P}\left(A_n \le -t\right) \le 2e^{-t^2/2},$$

since $\{A_n \ge t\} \cap \{A_n \le -t\} = \emptyset$, for $t > 0$.

(f) Show that the characteristic function $\phi_{A_n}(z) = \mathbb{E}\exp(izA_n)$ is given by

$$\phi_{A_n}(z) = \cos^n(z/\sqrt{n}).$$

Prove that $\lim_{n\to\infty} \phi_{A_n}(z) = \exp(-z^2/2)$.

**4 pts**

**ANS:** Setting $c = iz$ and applying $\cos\theta = (\exp(i\theta) + \exp(-i\theta))/2$ to the calculation of $\mathbb{E}\exp(cA_n)$ in part (c), we obtain the CF. Hence

$$\phi_{A_n}(z) = \left(1 - \frac{z^2}{2n} + o(1)\right)^n \to e^{-z^2/2},$$

as $n \to \infty$.

2. (a) [**Variant on standard problem.**] A program generates passwords $(X_1, X_2, \ldots, X_m)$ of length $m$, the component characters being chosen uniformly and independently from an alphabet of $N$ symbols. In other words, $\mathbb{P}(X_i = s_k) = 1/N$, for $1 \le k \le N$ and $1 \le i \le m$. We shall say that a password is *non-redundant* if its component characters are all different.

i. Show that the probability $p_m$ that a password of length $m$ is non-redundant is given by

$$p_m = \prod_{k=1}^{m-1} \left(1 - \frac{k}{N}\right).$$

**5 pts**

**ANS:** We have

$$\begin{aligned} p_m &= \frac{N(N-1)(N-2)\cdots(N-m+1)}{N^m} \\ &= \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)\cdots\left(1 - \frac{m-1}{N}\right), \end{aligned}$$

as required.

ii. Prove the inequality $1 - x \le \exp(-x)$, for $x \ge 0$. Hence prove that

$$p_m \le e^{-\frac{(m-1)m}{2N}}.$$

Furthermore, show that $p_m \le 10^{-q}$ if $m \ge c\sqrt{N}$, where $c = \sqrt{2q \ln 10}$. [Hint: One possible derivation of the inequality begins with the integral $\int_0^x \exp(-s)\, ds$. ]

**5 pts**

**ANS:** Using the hint, $x \ge 0$, and $\exp(-x) \le 1$, we obtain

$$x \ge \int_0^x e^{-s}\, ds = 1 - e^{-x},$$

and rearranging implies $\exp(-x) \ge 1 - x$, for $x \ge 0$. Hence

$$p_m \le \prod_{k=1}^{m-1} e^{-k/N} = e^{-\frac{1}{N}\sum_{k=1}^{m-1} k} = e^{\frac{-m(m-1)}{2N}},$$

using the elementary fact that $1 + 2 + \cdots + M = M(M+1)/2$. Thus, to achieve $p_m \le 10^{-q} = \exp(-q \ln 10)$, it's sufficient to choose $m$ so large that $\exp(-m(m-1)/(2N) \le \exp(-q \ln 10)$, i.e. $m(m-1) \ge 2Nq \ln 10$. This is true, with room to spare, if $(m-1)^2 \ge 2Nq$, i.e. if $m \ge 1 + \sqrt{2Nq}$. Thus the original question contained a typo! In the end, I decided to leave this typo uncorrected to teach you a valuable lesson: use logic and experiment, not faith in authority. I have, however, marked it generously.

(b) [**Variant on standard problem.**] Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with the exponential distribution at rate $\lambda$, that is, they share the probability density function

$$p_1(s) = \begin{cases} \lambda e^{-\lambda s}, & s \geq 0, \\ 0, & s < 0, \end{cases}$$

where $\lambda$ is a positive constant.

i. Let $p_n(s)$ be the probability density function for $X_1 + X_2 + \cdots + X_n$. Show that

$$p_n(s) = \begin{cases} \frac{\lambda^n s^{n-1} e^{-\lambda s}}{(n-1)!}, & s \geq 0, \\ 0, & s < 0. \end{cases}$$

**5 pts**

**ANS:** The students have seen two derivations of this result. Firstly, convolving the PDFs of $X_1 + \cdots + X_{n-1}$ and $X_1$,

$$p_n(s) = \int_{\mathbb{R}} p_{n-1}(y) p_1(s-y)\, dy = \int_0^s p_{n-1}(y) p_1(s-y)\, dy,$$

for $s \geq 0$, because the PDFs are nonzero if and only if $y \geq 0$ and $s - y \geq 0$. We can now proceed by induction, noting that the given formula is correct when $n = 1$. Assuming its validity for $n-1$, we obtain

$$\begin{aligned} p_n(s) &= \int_0^s \left( \frac{\lambda^{n-1} y^{n-2} e^{-\lambda y}}{(n-2)!} \right) \lambda e^{-\lambda(s-y)}\, dy \\ &= \frac{\lambda^n}{(n-2)!} e^{-\lambda s} \int_0^s y^{n-2}\, dy \\ &= \frac{\lambda^n}{(n-2)!} e^{-\lambda s} \left[ \frac{y^{n-1}}{n-1} \right]_0^s \\ &= \frac{\lambda^n}{(n-1)!} s^{n-1} e^{-\lambda s}. \end{aligned}$$

Alternatively, the student might observe that, since the random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n) \in \mathbb{R}^n$ has PDF

$$q(\mathbf{s}) = p_1(s_1) p_1(s_2) \cdots p_1(s_n), \qquad \mathbf{s} \in \mathbb{R}^n,$$

we obtain

$$\mathbb{P}\left(a \leq \mathbf{e}^T \mathbf{X} \leq b\right) = \int_{a \leq \mathbf{e}^T \mathbf{s} \leq b,\ \mathbf{s} \geq 0} \lambda^n e^{-\lambda \mathbf{e}^T \mathbf{s}}\, ds,$$

where $\mathbf{e} = (1, 1, \ldots, 1)^T \in \mathbb{R}^n$. Hence geometry implies the relation

$$\mathbb{P}\left(a \leq \mathbf{e}^T \mathbf{X} \leq b\right) = c_n \int_a^b \lambda^n e^{-\lambda u} u^{n-1}\, du,$$

for some constant $c_n$. To determine $c_n$, we set $a = 0$ and $b = \infty$, whence

$$1 = c_n \int_0^\infty \lambda^n e^{-\lambda u} u^{n-1}\, du = c_n \int_0^\infty e^{-v} v^{n-1}\, dv = c_n (n-1)!,$$

on setting $v = \lambda u$ and recalling the definition of the Gamma function.

ii. Show that the characteristic function $\phi_{X_1}(z) = \mathbb{E} \exp(izX_1)$ is given by

$$\phi_{X_1}(z) = \frac{\lambda}{\lambda - iz}.$$

Hence state the characteristic function for $X_1 + X_2 + \cdots + X_n$.

**5 pts**

**ANS:** We have

$$\begin{aligned}
\phi_{X_1}(z) &= \mathbb{E} e^{izX_1} \\
&= \int_0^\infty e^{izs} p_1(s)\, ds \\
&= \lambda \int_0^\infty e^{s(iz-\lambda)}\, ds \\
&= \lambda \left[ \frac{e^{s(iz-\lambda)}}{(iz-\lambda)} \right]_0^\infty \\
&= -\frac{\lambda}{iz - \lambda} \\
&= \frac{\lambda}{\lambda - iz}.
\end{aligned}$$

Finally,

$$\phi_{X_1+\cdots+X_n}(z) = (\phi_{X_1}(z))^n = \frac{\lambda^n}{(\lambda - iz)^n},$$

since CFs satisfy $\phi_{X+Y}(z) = \phi_X(z)\phi_Y(z)$ for independent random variables $X$ and $Y$.