8 Mathematical Methods for Computer Science (RJG)

(*a*) Given a random variable, X, with mean μ, variance σ^2 and a constant $c \geq 0$ prove *Chebyshev's inequality* in the form

$$\mathbb{P}(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$$

[5 marks]

(*b*) Suppose now that X is a random variable taking values in the interval $[a, b]$ with a mean μ and a variance σ^2. Define the function $f(\alpha) = \mathbb{E}((X - \alpha)^2)$ for $\alpha \in \mathbb{R}$ and show that $f(\alpha)$ is minimized by the choice $\alpha = \mu$. Show that

$$f\left(\frac{a+b}{2}\right) = \mathbb{E}((X-a)(X-b)) + \frac{(b-a)^2}{4}$$

and hence that $\text{Var}(X) \leq (b-a)^2/4$. In the case that X is a Bernoulli random variable show that $\text{Var}(X) \leq 1/4$. [5 marks]

(*c*) Let p be the fraction of computers that are running normally on some network and $1 - p$ the fraction that need rebooting. Suppose that you test n of the computers choosing independently and without replacement. Let X_i be the Bernoulli random variable recording the result of the ith test for $i = 1, \ldots, n$. Write $P_n = \sum_{i=1}^{n} X_i/n$ for the proportion of computers in your sample that were found to be running normally and show that

$$\mathbb{P}(|P_n - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2}$$

if p is known. However, if p is unknown show that

$$\mathbb{P}(|P_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}$$

[5 marks]

(*d*) Now suppose that you wish to determine the least sample size n such that

$$\mathbb{P}(|P_n - p| \geq \epsilon) \leq \delta$$

for given choices of ϵ and δ. What happens to the value of n as recommended by the Chebyshev inequality in part (*c*) in each of the following two cases?

(*i*) the value of ϵ is halved

(*ii*) the probability δ is halved

[5 marks]