Tutorial 4 Solutions

1. In this problem we assume that a single disk inside a RAID 5 array has died. In order to recompute the missing data, the data stored on all the remaining disks has to be read. The task is to consider the number of expected read errors during the reconstruction of the data. Let $N(b)$ be the number of read errors when trying to read b bits, which is assumed to be distributed according to a binomial distribution $B(b, p)$ with parameter value $p = 10^{-15}$. The resulting error process is a *Bernoulli process*, which can be considered the discrete-time version of a Poisson process. Note that p equals the probability of incorrectly reading a single bit. The model contains the assumption that bit read errors are independent for each bit (and thus for each hard disk). From standard literature we know that $E[N(b)] = b\,p$, which is the expected number of errors when reading b bits from any of the hard disks. Let us determine the expected number of errors when reading the complete contents of the K *other* disks in the RAID array. Due to linearity of expectation we know that

$$E[\sum_i N_i(b)] = \sum_i E[N_i(b)] = K\,b\,p,$$

where the index i runs from 1 to K. A different approach may rely on properties of the binomial distribution, which imply that the sum of K binomially distributed variables (with equal parameter p) again follows a binomial distribution. We obtain the following results.

Disk capacity	b	$E[N(b)]$
4 TB	$32 * 10^{12}$	0.032
16 TB	$128 * 10^{12}$	0.128
64 TB	$512 * 10^{12}$	0.512
256 TB	$2048 * 10^{12}$	2.048
1024 TB	$8192 * 10^{12}$	8.192

2. We now have $p = 10^{-16}$, which results in the following.

Disk capacity	b	$E[N(b)]$
4 TB	$32 * 10^{12}$	0.0032
16 TB	$128 * 10^{12}$	0.0128
64 TB	$512 * 10^{12}$	0.0512
256 TB	$2048 * 10^{12}$	0.2048
1024 TB	$8192 * 10^{12}$	0.8192