PROBLEM 1.

(a) Let the codeword length for the source letter u be $l(u) = -\log_2 \Pr(u)$. Since $\Pr(u)$ is an integer power of $1/2$, $l(u)$ is an integer. Furthermore,

$$\sum_u 2^{-l(u)} = \sum_u \Pr(u) = 1,$$

so $\{l(u)\}$ satisfies the Kraft inequality and

$$E[l(U)] = \sum_u \Pr(u) l(u) = -\sum_u \Pr(u) \log_2 \Pr(u) = H(U),$$

so that the expected codeword length is equal to the entropy of the source. Since Huffman coding has to do as well as any other coding scheme, and since no coding scheme can do better than the entropy, we conclude that the expected codeword length of the Huffman code is the same as the entropy.

(b) Since the source is stationary, the distribution of (X_{2k-1}, X_{2k}) is the same for any k. In method 1 we design a code for a source U whose distribution is identical to the distribution of (X_1, X_2), and since all probabilities of interest are integer powers of $1/2$, we see that the expected codeword length will be $H(X_1, X_2)$. Since the method encodes two letters at a time, the expected codeword length per letter is

$$\frac{1}{2} H(X_1, X_2).$$

(c) In method 2 we design $|\mathcal{X}|$ Huffman codes, one for each value of the preceding letter. When we are encoding X_k, if the preceding letter X_{k-1} is x' we use a Huffman code for the source that has distribution $p(x) = \Pr(X_k = x | X_{k-1} = x')$. By stationarity this is the same as $\Pr(X_2 = x | X_1 = x')$, and the Huffman code that is used when the preceding letter is x' has expected length

$$-\sum_x \Pr(X_2 = x | X_1 = x') \log_2 \Pr(X_2 = x | X_1 = x').$$

The probability that the preceding letter is x' is $\Pr(X_{k-1} = x') = \Pr(X_1 = x')$, and we see that the expected codeword length for method two is

$$\begin{aligned} -\sum_{x'} \Pr(X_1 = x') \sum_x \Pr(X_2 = x | X_1 = x') \log_2 \Pr(X_2 = x | X_1 = x') \\ = -\sum_{x,x'} \Pr(X_1 = x', X_2 = x) \log_2 \Pr(X_2 = x | X_1 = x') = H(X_2 | X_1). \end{aligned}$$

(d) Using the chain rule, that conditioning does not increase entropy, and stationarity we get:

$$H(X_1, X_2) = H(X_1) + H(X_2|X_1) \geq H(X_1|X_0) + H(X_2|X_1) = 2H(X_2|X_1).$$

Thus, the expected codeword length of method 1, $\frac{1}{2}H(X_1, X_2)$, is at least as large as $H(X_2|X_1)$, the expected codeword length of method 2. So, method 2 is better.

PROBLEM 2.

(a) We have

$$I(X;Y) = h(Y) - h(Y|X) = h(Y) - h(Z|X) = h(Y) - h(Z).$$

where the last equality is because Z is independent of X.

(b) In the natural log basis,

$$h(Z) = -\int f_Z(z) \ln f_Z(z)\, dz = \int_0^\infty z e^{-z}\, dz = 1 \text{ nats}.$$

(c) Since $Y = X + Z$, the expectation of Y, $E[Y]$ equals $E[X] + E[Z]$. Since $E[X]$ is constrained to be less than or equal to P and $E[Z] = 1$, we see that $E[Y] \leq P + 1$. Since X is constrained to be non-negative and so is Z, we see that Y is also constrained to be non-negative.

(d) From homework 10, problem 5 we know that among non-negative random variables of a given expectation λ, the one with density $p(y) = e^{-y/\lambda}/\lambda$ has the largest differential entropy. [Proof: for distribution q with mean λ,

$$\begin{aligned}
0 &\leq D(q\|p) \\
&= \int q(y) \log q(y)\, dy - \int q(y) \log p(y)\, dy \\
&= \int q(y) \log q(y)\, dy - \int p(y) \log p(y)\, dy \\
&= h(p) - h(q)
\end{aligned}$$

where the second equality is because $\log p(y)$ is of the form $\alpha + \beta y$, and has the same expectation under both p and q.) This maximal entropy in natural units is

$$\int_0^\infty \frac{e^{-y/\lambda}}{\lambda} [\ln \lambda + y/\lambda]\, dy = \ln \lambda + 1 \text{ nats}.$$

Thus, the differential entropy of Y is less than $1 + \ln E[Y] \leq 1 + \ln(1 + P)$, which implies

$$C \leq \ln(1 + P) \text{ nats}$$

At this point, we do not know if Y can be made to have an exponential distribution with mean $1 + P$ so we cannot know if this above inequality is an equality of not.

(e) For the given density of X

$$f_Y(y) = \int_{-\infty}^{\infty} f_Z(z) f_X(y-z)\, dz$$

Note that $f_Z(z) = 0$ for $z < 0$ and $f_X(y-z) = 0$ when $z > y$. Thus, if $y < 0$ the integrand is zero, and when $y \geq 0$ the integrand is non-zero only when $0 \leq z \leq y$. Thus, for $y \geq 0$

$$\begin{aligned} f_Y(y) &= \int_0^y e^{-z}[\mu\delta(y-z) + (1-\mu)\mu e^{\mu(z-y)}]\, dz \\ &= \mu e^{-y} + \mu e^{-\mu y} \int_0^y (1-\mu) e^{(\mu-1)z}\, dz \\ &= \mu e^{-y} - \mu e^{-\mu y} e^{(\mu-1)z}\Big|_{z=0}^{z=y} \\ &= \mu e^{-y} - \mu[e^{-y} - e^{-\mu y}] \\ &= \mu e^{-\mu y}. \end{aligned}$$

(f) By taking $\mu = 1/(1+P)$, we see that there is a density on X which makes the the density of Y an exponential with mean $1+P$. Furthermore, this density on X makes X non-negative, and, $E[X] = E[Y] - E[Z] = P$. Thus, the bound on part (d) can be achieved.

PROBLEM 3.

(a) The channel is a binary symmetric channel with crossover probability ϵ. Hence, the capacity is achieved by an input X taking the values 0 and 1 with probability 1/2, and the capacity is $\log 2 + \epsilon \log \epsilon + (1-\epsilon)\log(1-\epsilon)$.

(b) The distance of the received sequence $\mathbf{y}$ from the transmitted codeword is given by $\sum_{i=1}^n Z_i = \sum_{i=1}^d Z_i + \sum_{i>d} Z_i$. The distance of the received sequence from the other codeword is given by $\sum_{i=1}^d (1-Z_i) + \sum_{i>d} Z_i$. Thus, the received sequence will be closer to the wrong codeword only if

$$\sum_{i=1}^d Z_i \geq d/2.$$

(c) The Chernoff bound applies to our expression for the error probability with $\alpha = 1/2$. Choosing $e^s = (1-\epsilon)/\epsilon$ (which is larger than 1 because $\epsilon < 1/2$ which makes sure that $s > 0$), we see that

$$\begin{aligned} \Pr(\text{error}) &\leq \left(e^{-s/2} E[e^{sZ}]\right)^d \\ &= \left(\frac{\epsilon^{1/2}}{(1-\epsilon)^{1/2}}\left[(1-\epsilon) + \epsilon\frac{1-\epsilon}{\epsilon}\right]\right)^d \\ &= \left(2\sqrt{\epsilon(1-\epsilon)}\right)^d \\ &= \left[4\epsilon(1-\epsilon)\right]^{d/2}. \end{aligned}$$

(d) We know that the error probability is given by

$$\begin{aligned}
\Pr\Big(\sum_{i=1}^{d} Z_i \geq d/2\Big) &= \sum_{k=(d+1)/2}^{d} \Pr\Big(\sum_{i=1}^{d} Z_i = k\Big) \\
&= \sum_{k=(d+1)/2}^{d} \binom{d}{k} \epsilon^k (1-\epsilon)^{d-k}.
\end{aligned}$$

Keeping only the first term in this sum we get the required lower bound.

(e) When the codewords are chosen randomly, the distance between them is given by

$$D = \sum_{i=1}^{n} \mathbf{1}\{U_i \neq V_i\}$$

where $\mathbf{1}(E)$ is a function that equals 1 is E is true and 0 is E is false. Thus,

$$E[D] = \sum_{i=1}^{n} E[\mathbf{1}\{U_i \neq V_i\}] = \sum_{i=1}^{n} \Pr(U_i \neq V_i) = n/2,$$

since for each i, $U_i \neq V_i$ with probability 1/2.