Problem 1.

(a) We have

\[
\Pr[U(1) \neq U^n \mid U^n = u^n] = \Pr[U(1) \neq u^n \mid U^n = u^n] = \mathcal{Y}^n_i \Pr[U(1) = u_i] = 1 - \Pr[U(1) = u^n] = 1 - \prod_{i=1}^{\mathcal{Y}^n_i} p_U(u_i),
\]

where \((*)\) follows from the independence of \(U(1)\) and \(U^n\).

(b) An encoding failure happens if and only if \(U(m) \neq U^n\) for every \(m = 1, 2, \ldots, M\). Therefore,

\[
\Pr[\text{"failure"} \mid U^n = u^n] = \Pr[U(m) \neq U^n, \forall m = 1, \ldots, M \mid U^n = u^n] \\
= \Pr[U(m) \neq u^n, \forall m = 1, \ldots, M \mid U^n = u^n] \\
= \prod_{m=1}^{\mathcal{Y}^n_i} 1 - p_U(u_i) = 1 - \prod_{i=1}^{\mathcal{Y}^n_i} p_U(u_i)
\]

(c) Note that if \(u^n \in T^n_\epsilon(p_U)\), then 
\(\sum_{i=1}^{\mathcal{Y}^n_i} p_U(u_i) \geq 2^{-nH(U)(1+\epsilon)}\), which implies

\[
\Pr[\text{"failure"} \mid U^n = u^n] = 1 - \prod_{i=1}^{\mathcal{Y}^n_i} p_U(u_i) \leq 1 - 2^{-nH(U)(1+\epsilon)} M
\]

\[\leq \exp -M2^{-nH(U)(1+\epsilon)} = \exp -2^{nR-nH(U)(1+\epsilon)} .\]

where \((*)\) follows from the hint. Therefore, we have

\[
\Pr[\text{"failure"} \mid U^n \in T^n_\epsilon(p_U)] = \frac{\Pr[\text{"failure"}, U^n \in T^n_\epsilon(p_U)]}{\Pr[U^n \in T^n_\epsilon(p_U)]} \\
= \frac{\sum_{u^n \in T^n_\epsilon(p_U)} \Pr[\text{"failure"}, U^n = u^n]}{\Pr[U^n \in T^n_\epsilon(p_U)]} \\
= \frac{\Pr[U^n \in T^n_\epsilon(p_U)]}{\Pr[U^n \in T^n_\epsilon(p_U)]} \exp -2^{nR-nH(U)(1+\epsilon)} \Pr[U^n = u^n] \\
\leq \exp -2^{nR-nH(U)(1+\epsilon)} \frac{\sum_{u^n \in T^n_\epsilon(p_U)} \Pr[U^n = u^n]}{\Pr[U^n \in T^n_\epsilon(p_U)]} \\
= \exp -2^{nR-nH(U)(1+\epsilon)} \frac{\Pr[U^n \in T^n_\epsilon(p_U)]}{\Pr[U^n \in T^n_\epsilon(p_U)]} \\
= \exp -2^{nR-nH(U)(1+\epsilon)} .
\]
(d) Assume \( R > H(U) \), then there exists \( \epsilon > 0 \) such that \( R > H(U)(1 + \epsilon) \). We have

\[
\Pr[\text{“failure”}] = \Pr[\text{“failure”, } U^n \in T^n_e(p_U)] + \Pr[\text{“failure”, } U^n \notin T^n_e(p_U)]
\]

\[
= \Pr[\text{“failure” } | U^n \in T^n_e(p_U)] \Pr[U^n \in T^n_e(p_U)] + \Pr[\text{“failure” } , U^n \notin T^n_e(p_U)]
\]

\[
\leq \Pr[\text{“failure” } | U^n \in T^n_e(p_U)] + \Pr[\text{“failure” } , U^n \notin T^n_e(p_U)]
\]

\[
\leq \exp -2^{nR-nH(U)(1+\epsilon)} + \Pr[U^n \notin T^n_e(p_U)].
\]

Since \( R > H(U)(1 + \epsilon) \) both terms in the above go to 0 as \( n \to \infty \). Hence, \( \Pr[\text{“failure”}] \to 0 \) as \( n \) gets large.

**Problem 2.** Let the input distribution be \( p \). We thus have

\[
p(-1) + p(0) + p(1) = 1 \quad p(-1) \geq 0, p(0) \geq 0, p(1) \geq 0
\]

(since \( p \) is a distribution) and, to satisfy \( E[b(X)] \leq \beta \) we must have

\[
p(-1) + p(1) = 1 - p(0) \leq \beta.
\]

Moreover,

\[
I(X;Y) = H(Y) - H(Y|X)
\]

\[
\equiv H(Y) - p(0)
\]

\[
\leq 1 - p(0)
\]

\[
\leq \max \{1, \beta\}.
\]

where (a) follows because given \( \{X = -1\} \) or \( \{X = 1\} \) there is no uncertainty in \( Y \) while given \( \{X = 0\} \), \( Y \) is uniformly distributed in \( \{-1, 1\} \), (b) holds since \( Y \) is binary with equality if \( p(-1) + \frac{1}{2}p(0) = p(1) + \frac{1}{2}p(0) = \frac{1}{2} \) (which happens if we choose \( p(1) = p(-1) = \frac{1}{2}(1 - p(0)) \)) and (c) holds because of the cost constraint and is equality if we choose \( p(0) = \max \{1 - \beta, 0\} \). Hence, the capacity is

\[
C = \begin{cases} 
\beta, & \text{if } \beta \leq 1 \\
1, & \text{if } \beta > 1. 
\end{cases}
\]

**Problem 3.**

\[
h(X) = \frac{1}{2} \log(2\pi e\sigma^2_x)
\]

\[
h(Y) = \frac{1}{2} \log(2\pi e\sigma^2_y)
\]

\[
h(X,Y) = \frac{1}{2} \log((2\pi e)^2 \det(K)) = \frac{1}{2} \log((2\pi e)^2(\sigma^2_x\sigma^2_y - \rho^2\sigma^2_x\sigma^2_y))
\]

\[
I(X,Y) = h(X) + h(Y) - h(X,Y) = \frac{1}{2} \log \frac{1}{1-\rho^2}
\]

Note that the result does not depend on \( \sigma_x, \sigma_y \), which says that normalization does not change the mutual information.

**Problem 4.**

(a) All rates less than \( \frac{1}{2} \log(1 + \frac{P}{\sigma}) \) are achievable.
(b) The new noise $Z_1 - \rho Z_2$ has zero mean and variance $E((Z_1 - \rho Z_2)^2) = \sigma^2(1 - \rho^2)$. Therefore, all rates less than $\frac{1}{2} \log_2(1 + \frac{P}{\sigma^2(1 - \rho^2)})$ are achievable.

(c) The capacity is $C = \max I(X;Y_1,Y_2) = \max(h(Y_1, Y_2) - h(Z_1, Z_2)) = \frac{1}{2} \log_2(1 + \frac{P}{\sigma^2(1 - \rho^2)})$. This shows that the scheme used in (b) is a way to achieve capacity.

**Problem 5.**

(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.

From Homework 9, Problem 7 we know that among non-negative random variables of a given expectation $\lambda$, the one with density $p(y) = e^{-y/\lambda}/\lambda$ has the largest differential entropy. This differential entropy in natural units is

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

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 or not.

(d) The Laplace transform of the random variable $Y$ is $E(e^{sY}) = E(e^{s(X+Z)}) = E(e^{sX})E(e^{sZ})$, where the latter equality follows from the independence of $X$ and $Z$. Therefore we have that $E(e^{sX}) = \frac{E(e^{sY})}{E(e^{sZ})}$. Computing $E(e^{sY})$,

$$E(e^{sY}) = \int_0^\infty e^{sy} f_Y(y) \, dy$$

$$= \int_0^\infty e^{sy} \mu e^{-\mu y} \, dy$$

$$= \frac{\mu}{\mu - s} \quad \forall s \leq \mu.$$

The expectation is not defined for $s > \mu$ (as the integral blows up). Likewise, we evaluate $E(e^{sZ}) = \frac{1}{1-s}$ (defined for $s \leq 1$). Therefore for $s \leq \min(1, \mu)$, we can
evaluate $E(e^{sX})$ as
\[
E(e^{sX}) = \frac{E(e^{sY})}{E(e^{sZ})} = \frac{1 - s}{\mu - s} = \mu + (1 - \mu) \frac{\mu}{\mu - s}
\]

Inverting the Laplace transform $E(e^{sX})$ gives us the distribution of the $X$ that gives an exponential distribution for $Y$. From inspection, we can deduce this distribution of $X$ to be
\[
f_X(x) = \mu \delta(x) + (1 - \mu) \mu e^{-\mu x} \quad x \geq 0
\]

Notice that the distribution is a convex combination of the exponential distribution and the distribution that puts all the mass on one point (in this case the point $x = 0$).

(e) By taking $\mu = 1/(1 + P)$, we see that there is a density on $X$ which makes 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 of part (c) can be achieved.

**Problem 6.**

(a)
\[
F(p, r_p) - F(p, r) = \sum_{x \in X} \sum_{y \in Y} p(x) P(y|x) \log_2 \frac{r_p(x|y)}{r(x|y)}
= \sum_{x \in X} \sum_{y \in Y} p(x) P(y|x) \log_2 \frac{P(x)P(y|x)}{P(x')P(y|x')} \quad x' \in X
= D(P_1 \parallel P_2) \geq 0,
\]

where $P_1(x, y) := p(x)P(y|x)$ and $P_2(x, y) := r(x|y) \sum_{x' \in X} p(x')P(y|x')$.

(b) We can rewrite $F(p, r)$ as follows:
\[
F(p, r) = \sum_{x \in X} p(x) P(y|x) \log_2 r(x|y) + \sum_{x \in X} p(x) \log_2 \frac{1}{p(x)}.
\]

The first term in (1) is linear in $p$ while the second term is strictly concave in $p$ (since the function $t \rightarrow t \log_2 \frac{1}{t}$ is strictly concave). Therefore, $F(p, r)$ is strictly concave in $p$.

The first term in 1 is concave in $r$ (since the function $\log_2$ is concave) and the second term is constant with respect to $r$. Therefore, $F(p, r)$ is concave in $r$.

(c) For every $x \in X$, we have:
\[
\frac{\partial F(p, r_k)}{\partial p(x)} = \sum_{y \in Y} P(y|x) \log_2 r_k(x|y) + \log_2 \frac{1}{p(x)} - \frac{1}{\ln 2}.
\]
A probability distribution $p$ satisfies the Kuhn-Tucker conditions if and only if there exists a real number $\lambda$ such that for all $x \in \mathcal{X}$, we have $\frac{\partial F(p, r_k)}{\partial p(x)} \leq \lambda$ with equality if $p(x) > 0$. Therefore, for all $x \in \mathcal{X}$ we have:

$$X \left( y \in \mathcal{Y} \right) \log_2 r_k(x|y) - \log_2(p(x)) \leq \lambda',$$

where $\lambda' = \lambda + \frac{1}{\ln 2}$. This shows that $p(x) \geq 2^{-\lambda'} \alpha_k(x)$.

If $p(x) > 0$, we have $p(x) = 2^{-\lambda'} \alpha_k(x)$, and if $p(x) = 0$ we must also have $p(x) = 2^{-\lambda'} \alpha_k(x) = 0$ since $2^{-\lambda'} 2\sum_{y \in \mathcal{Y}} p(y|x) \log_2 r_k(x|y) \geq 0$. We conclude that $p(x) = 2^{-\lambda'} \alpha_k(x)$ in all cases. Therefore, $1 = 2^{-\lambda'} \mathbb{P}_x \alpha_k(x)$, and $\lambda' = \log_2 \alpha_k(x)$. We conclude that the only distribution that satisfies the Kuhn-Tucker conditions is the one given by $p(x) = \mathbb{P}_{x' \in \mathcal{X}} \alpha_k(x')$. On the other hand, the fact that $F(p, r_k)$ is concave in $p$ shows that it admits a maximum $p_{k+1}$, which has to satisfy the Kuhn-Tucker conditions. Therefore, $p_{k+1}(x) = \mathbb{P}_{x' \in \mathcal{X}} \alpha_k(x')$.

(d) $C \geq F(p_{k+1}, r_{k+1})$ since $F(p_{k+1}, r_{k+1}) = I(X; Y)|_{p_X = p_{k+1}}$. This implies that $C \geq F(p_{k+1}, r_k)$ since $F(p_{k+1}, r_{k+1}) \geq F(p_{k+1}, r_k)$. On the other hand, we have

$$F(p_{k+1}, r_k) = \mathbb{E}_{x \in \mathcal{X}} \mathbb{P}_{x' \in \mathcal{X}} \alpha_k(x') \sum_{y \in \mathcal{Y}} \frac{\mathbb{P}(y|x) \log_2 r_k(x|y)}{\alpha_k(x)} = \mathbb{E}_{x \in \mathcal{X}} \mathbb{P}_{x' \in \mathcal{X}} \alpha_k(x') \log_2(\alpha_k(x)) \sum_{y \in \mathcal{Y}} \frac{P(y|x)}{p_k(x)} = \mathbb{X}_{y \in \mathcal{Y}} \log_2 \frac{r_k(x|y)}{p_k(x)}.$$

(e)

$$\log_2 \frac{\alpha_k(x)}{p_k(x)} = \log_2 \alpha_k(x) - \log_2 p_k(x) = \mathbb{X}_{y \in \mathcal{Y}} \log_2 \frac{r_k(x|y)}{p_k(x)} = \mathbb{X}_{y \in \mathcal{Y}} \log_2 \frac{P(y|x)}{p_k(x)}.$$

(f) Given that $\log_2 \frac{\alpha_k(x)}{p_k(x)} = \mathbb{X}_{y \in \mathcal{Y}} \log_2 \frac{P(y|x)}{p_k(x)}$, the inequality $C \leq \mathbb{X}_{x \in \mathcal{X}} \left| p^*(x) \log_2 \frac{\alpha_k(x)}{p_k(x)} \right|$ is a direct application of Homework 8 Problem 5.
(g) From (d) and (f), we have:

\[
C - F(p_{k+1}, r_k) \\
\leq \sum_{x \in \mathcal{X}} p^*(x) \log_2 \frac{\alpha_k(x)}{p_k(x)} - \log_2 \alpha_k(x) = \sum_{x \in \mathcal{X}} p^*(x) \log_2 \frac{\alpha_k(x)}{p_k(x)} - \log_2 \alpha_k(x')
\]

\[
= \sum_{x \in \mathcal{X}} p^*(x) \log_2 \frac{\alpha_k(x)}{p_k(x)} \frac{p_{k+1}(x)}{p_{k+1}(x')} = \sum_{x \in \mathcal{X}} p^*(x) \log_2 \frac{p_{k+1}(x)}{p_k(x)} \leq \max_{x \in \mathcal{X}} \log_2 \frac{p_{k+1}(x)}{p_k(x)}.
\]

(h) We prove it by induction on \( n \). The result is trivial for \( n = 0 \). Now assume that it is true for \( n \), and let us prove it for \( n + 1 \):

\[
\sum_{k=0}^{n+1} (C - F(p_{k+1}, r_k)) = C - F(p_{n+2}, r_{n+1}) + \sum_{k=0}^{n} (C - F(p_{k+1}, r_k))
\]

\[
\leq \sum_{x \in \mathcal{X}} p^*(x) \log_2 \frac{p_{n+2}(x)}{p_{n+1}(x)} + \sum_{x \in \mathcal{X}} p^*(x) \log_2 \frac{p_{n+1}(x)}{p_0(x)}
\]

\[
= \sum_{x \in \mathcal{X}} p^*(x) \log_2 \frac{p_{n+2}(x)}{p_0(x)}.
\]

On the other hand, since \( p_{n+1}(x) \leq 1 \) for all \( x \in \mathcal{X} \), we have:

\[
\sum_{x \in \mathcal{X}} p^*(x) \log_2 \frac{p_{n+1}(x)}{p_0(x)} \leq \sum_{x \in \mathcal{X}} p^*(x) \log_2 \frac{1}{|\mathcal{X}|} = \log_2 |\mathcal{X}|.
\]

(i) The sequence \( s_n = \sum_{k=0}^{n} (C - F(p_{k+1}, r_k)) \) is increasing and upper-bounded, thus convergent, which implies that the sequence \( C - F(p_{k+1}, r_k) = s_k - s_{k-1} \) converges to zero. Therefore, \( F(p_{k+1}, r_k) \) converges to \( C \).