Problem 1 [0 points]

Let $F$ be a finite set of $n$ elements. We have that $C = \{A: A \in \mathcal{A} \text{ or } A \in \mathcal{B}\}$, so if $C$ picks out $G \subseteq F$, then either $\mathcal{A}$ picks out $G$ or $\mathcal{B}$ picks out $G$. Thus, for any $F \in \mathcal{F}_n$, the total number of subsets picked out by $C$ is the total number of distinct $G \subseteq F$ picked out by either $\mathcal{A}$ or $\mathcal{B}$. Therefore, for any finite set $F$:

$$S(C, F) \leq S(\mathcal{A}, F) + S(\mathcal{B}, F)$$

Taking the supremum over all $F \in \mathcal{F}_n$ on both sides:

$$s_n(C) = \sup_{F \in \mathcal{F}_n} S(C, F) \leq \sup_{F \in \mathcal{F}_n} (S(\mathcal{A}, F) + S(\mathcal{B}, F))$$

Problem 2 [25 points]

We have that $C = \{A \cup B: A \in \mathcal{A}, B \in \mathcal{B}\}$. Let $F \in \mathcal{F}_n$, then for $C \in C$,

$$C \cap F = (A \cup B) \cap F = (A \cap F) \cup (B \cap F)$$

Let $m_A, m_B$ be the number of subsets of $F$ that $\mathcal{A}$ and $\mathcal{B}$ can pick out. Then the number of distinct sets of the form $(A \cup B) \cap F$ is the total number of distinct unions of the form $(A \cap F) \cup (B \cap F)$ which is bounded by $m_A m_B$.

Let $m_A, m_B$ be the number of subsets of $F$ that $\mathcal{A}$ and $\mathcal{B}$ can pick out. Then the number of distinct sets of the form $(A \cup B) \cap F$ is the total number of distinct unions of the form $(A \cap F) \cup (B \cap F)$ which is bounded by $m_A m_B$. Thus:

$$S(C, F) \leq S(\mathcal{A}, F) S(\mathcal{B}, F)$$

---

1Formal proof: Let $\mathcal{A}_F = \{A \cap F: A \in \mathcal{A}\}$, $\mathcal{B}_F = \{B \cap F: B \in \mathcal{B}\}$, and $\mathcal{C}_F = \{(A \cap F) \cup (B \cap F): A \in \mathcal{A}, B \in \mathcal{B}\}$. Define a map $\Phi: \mathcal{A}_F \times \mathcal{B}_F \to \mathcal{C}_F$ by $\Phi((A \cap F) \times (B \cap F)) = (A \cap F) \cup (B \cap F)$. Then $\Phi$ is surjective, so $|\mathcal{C}_F| \leq |\mathcal{A}_F| \times |\mathcal{B}_F| = m_A m_B$, i.e. $|\{(A \cap F) \cup (B \cap F): A \in \mathcal{A}, B \in \mathcal{B}\}| \leq m_A m_B$, where $|\mathcal{X}|$ means number of elements of $\mathcal{X}$ for any set $\mathcal{X}$. 
Again taking the supremum over all $F \in \mathcal{F}_n$ on both sides:

\[
s_n(C) = \sup_{F \in \mathcal{F}_n} S(C, F) \leq \sup_{F \in \mathcal{F}_n} (S(A, F) S(B, F))
\]

\[
\leq \sup_{F \in \mathcal{F}_n} S(A, F) \sup_{F \in \mathcal{F}_n} S(B, F) = s_n(A) s_n(B)
\]

**Problem 3 [25 points]**

First of all note that if $f : \mathbb{R} \to \mathbb{R}$ with $\sup_x f(x) \leq C < \infty$, we have $L_1 \supseteq L_2$, for any $t_1 \leq t_2$, so that $L_0$ is the largest possible set in $\mathcal{A}$, while $L_C = \emptyset$. We have to distinguish 2 cases:

- “$f(x) \equiv 0$”. In this case $L_t = \emptyset$ for all $t \geq 0$. So for any finite set $F = \{a\} \in \mathcal{F}_1$ with one element, $L_t \cap F = \emptyset$ always, i.e. $\mathcal{A}$ can only pick out $\emptyset$. So $S(\mathcal{A}, F) = 1$ for any $F \in \mathcal{F}_1$, and $s_1(\mathcal{A}) = \sup_{\mathcal{F}_1} S(\mathcal{A}, F) = 1 < 2^1$. Thus $\text{VC} < 1$. On the other hand, finite set with zero element is only $\emptyset$ (i.e. $\mathcal{F}_0 = \{\emptyset\}$), and $\mathcal{A}$ can always pick out $\emptyset$. So $s_0(\mathcal{A}) = S(\mathcal{A}, \emptyset) = 0 = 2^0$, so $\text{VC} \geq 0$. From these, $\text{VC} = 0$.

- “$f$ such that $\exists a \in \mathbb{R}$ with $f(a) > 0$, i.e. $L_0 \neq \emptyset$”. Let $F = \{a\} \in \mathcal{F}_1$, with $a \in L_0$. Then $L_0 \cap F = \{a\}$ and $L_{f(a)} \cap F = \emptyset$, i.e. $\mathcal{A}$ can pick out both $\{a\}$ and $\emptyset$ from $F$ so that $S(\mathcal{A}, F) = 2$. Thus $s_1(\mathcal{A}) = \sup_{\mathcal{F}_1} S(\mathcal{A}, F) = 2 = 2^1$, which means $\text{VC} \geq 1$. Now, let $F = \{a_1, a_2\} \in \mathcal{F}_2$ be any finite set with two elements with $a_1, a_2 \in L_0$, and without loss of generality, assume $f(a_1) \geq f(a_2)$. If $a_2 \in L_t$ for some $t \geq 0$, then $f(a_2) > t$, which implies $f(a_1) \geq f(a_2) > t$, so $a_1 \in L_t$ as well. Hence it is impossible to have $L_t \cap F = \{a_1\}$ for any $t$, i.e. $\mathcal{A}$ cannot pick out $\{a_1\}$. In the case where $F = \{a_1, a_2\}$ with $a_i \notin L_0$ for some $i = 1, 2$, $\mathcal{A}$ could never pick out $\{a_i\}$. So $S(\mathcal{A}, F) \leq 3$ for any $F \in \mathcal{F}_2$, and $s_2(\mathcal{A}) = \sup_{\mathcal{F}_2} S(\mathcal{A}, F) = 3 < 2^2$. So $\text{VC} < 2$. From these, $\text{VC} = 1$.

**Remark**

What if $L_t = \{x : f(x) \geq t\}$ instead of $L_t = \{x : f(x) > t\}$? We would have that for any $a \in \mathbb{R}$, $L_0 \cap \{a\} = \mathbb{R} \cap \{a\} = \{a\}$ and $L_{f(a)+\varepsilon} \cap \{a\} = \emptyset$, for any $\varepsilon > 0$, so $\mathcal{A}$ can always pick out both $\{a\}$ and $\emptyset$. The rest of the proof is same as in the second case of the original proof. So $\text{VC} = 1$ for any $f : \mathbb{R} \to \mathbb{R}$ with $\sup_x f(x) \leq C < \infty$, even the case $f(x) \equiv 0$. 

2
Example: $L_{0.6} \subseteq L_{0.4}$ so that, for instance, $\{7\}$ cannot be picked out from set $F = \{4, 7\}$.

**Problem 4: Wasserman 5.2 [0 points]**

Let $\mathbb{E}[X_n] = \mu_n$. We can decompose

\[
\mathbb{E}[(X_n - b)^2] = \mathbb{E}[(X_n - \mu_n + \mu_n - b)^2] \\
= \mathbb{V}(X_n) + 2(\mu_n - b)\mathbb{E}[(X_n - \mu_n)] + (\mu_n - b)^2 \\
= \mathbb{V}(X_n) + (\mu_n - b)^2
\]

Thus, if $\mathbb{V}(X_n) \to 0$ and $\mu_n \to b$, then $\mathbb{E}[(X_n - b)^2] \to 0$. Since all terms in the equation above are nonnegative, we also have that $\mathbb{E}[(X_n - b)^2] \to 0$ only if both $\mathbb{V}(X_n) \to 0$ and $(\mu_n - b)^2 \to 0$, and thus $\mu_n \to b$.

**Remark**

Note that this result holds for $b$ constant, so non-random. Otherwise if $X_n \overset{q,m.}{\to} X$ with $X$ random variable, then $V(X) > 0 \Rightarrow V(X_n) \to V(X) > 0$. 


Problem 5: Wasserman 5.4 [25 points]

The distribution of $X_n$ tells us that $X_n$ is equal to $1/n$ with probability increasing in $n$. This suggests that $X_n$ might converge to zero, or at least to a constant $c$. In fact, since $X_n > 0$, by Markov’s inequality, for any $\varepsilon > 0$

$$P(X_n > \varepsilon) \leq \frac{\mathbb{E}[X_n]}{\varepsilon} = \frac{1}{n} \left(1 - \frac{1}{n^2}\right) + \frac{1}{n} \to 0$$

(1)

i.e. $X_n \xrightarrow{P} 0$. We could have also computed for a generic (nonnegative) constant $c, P(|X_n - c| > \varepsilon) \leq \frac{\mathbb{E}[|X_n - c|]}{\varepsilon} \to |c|/\varepsilon$, which is in fact zero only for $c = 0$.

Now, if also $X_n$ converged in q.m. to something, that would have to be 0, otherwise if $X_n \xrightarrow{q.m.} W$, where $W \neq 0$ or $W$ random variable, then we would have $X_n \xrightarrow{q.m.} W \Rightarrow X_n \xrightarrow{P} W$, which would be a contradiction. Thus we have to check if $X_n \xrightarrow{q.m.} 0$. By using the result of Problem 4 (Wasserman 5.2) for $b = 0$, we find that

$$V(X_n) = \frac{1}{n^2} \left(1 - \frac{1}{n^2}\right) + 1 - \left[\frac{1}{n} \left(1 - \frac{1}{n^2}\right) + \frac{1}{n}\right]^2 \to 1$$

(2)

so that $X_n$ does not converge in quadratic mean (to anything).

Problem 6: Wasserman 5.7 [0 points]

We just have to prove $(b)$, since it easily implies $(a)$. In fact $\mathbb{P}(X_n > \varepsilon) \leq \mathbb{P}(Y_n > \varepsilon)$. Since $X_n \sim \text{Poisson}(1/n), \forall \varepsilon > 0$

$$\mathbb{P}(Y_n > \varepsilon) = \mathbb{P} \left( X_n > \frac{\varepsilon}{n} \right) \leq \mathbb{P}(X_n > 0) = 1 - \mathbb{P}(X_n = 0) = 1 - e^{-\frac{1}{n}} \to 0$$

Alternate solution. Notice that, we can easily prove the stronger statement $Y_n^{(k)} = n^k X_n \xrightarrow{P} 0, \forall k \in \mathbb{N}$, such that cases (a) and (b) are obtained for $k = 0, 1$, respectively.
Problem 7: Wasserman 5.9 [25 points]

It is convenient to try to prove the statements from the strongest to the weakest, because $X_n \xrightarrow{a.m.} X \Rightarrow X_n \xrightarrow{p} X \Rightarrow X_n \rightsquigarrow X$. However, we will see that $X_n \not\xrightarrow{p} X$, but $X_n \xrightarrow{p} X \Rightarrow X_n \rightsquigarrow X$.

- **Convergence in quadratic mean**

\[
\mathbb{E}[(X_n - X)^2] = \mathbb{E}[(X_n - X)^2|X_n = X]P(X_n = X) + \mathbb{E}[(X_n - X)^2|X_n = e^n]P(X_n = e^n)
\]
\[
\geq \mathbb{E}[(X_n - X)^2|X_n = e^n]P(X_n = e^n)
\]
\[
= \mathbb{E}[(e^n - X)^2|X_n = e^n] \frac{1}{n}
\]
\[
\geq \mathbb{E}[(e^n - 1)^2|X_n = e^n] \frac{1}{n} \quad \text{[since $|e^n - X| \geq e^n - 1$]}
\]
\[
\geq \frac{(e^n - 1)^2}{n} \xrightarrow{n \to \infty} \infty.
\]

or alternatively

\[
\mathbb{E}[(X_n - X)^2] \geq \mathbb{E}[(X_n - X)^2 I(X_n = e^n)] = \mathbb{E}[(e^n - X)^2 I(X_n = e^n)]
\]
\[
\geq \mathbb{E}[(e^n - 1)^2 I(X_n = e^n)] \quad \text{[since $|e^n - X| \geq e^n - 1$]}
\]
\[
= (e^n - 1)^2 P(X_n = e^n) = \frac{(e^n - 1)^2}{n} \xrightarrow{n \to \infty} \infty.
\]

i.e. $X_n$ does not converge to $X$ in quadratic mean.

**Alternate solution explicitly assuming hierarchical structure.** By explicitly assuming a hierarchical structure where $P(X = -1) = P(X = 1) = 1/2$ and conditionally on $X$, we have that $X_n = X$ with probability $1 - 1/n$, and $X_n = e^n$ with probability $1/n$, then the two following proofs are also possible.

a) Note that \(\mathbb{E}[(X_n - X)^2] = \mathbb{E}[\mathbb{E}[(X_n - X)^2|X]]\). Thus:

\[
\mathbb{E}[(X_n - X)^2|X] = \mathbb{E}[(X_n - X)^2|X, X_n = X]P(X_n = X|X) + \mathbb{E}[(X_n - X)^2|X, X_n = e^n]P(X_n = e^n|X)
\]
\[
= (X - X)^2 \left(1 - \frac{1}{n}\right) + (e^n - X)^2 \frac{1}{n}
\]
\[
= (e^n - X)^2 \frac{1}{n}
\]

Therefore

\[
\mathbb{E}[\mathbb{E}[(X_n - X)^2|X]] = \frac{1}{n} (e^n - 1)^2 P(X = 1) + \frac{1}{n} (e^n + 1)^2 P(X = -1)
\]
\[
= \frac{1}{2n} (e^n - 1)^2 + \frac{1}{2n} (e^n + 1)^2 \xrightarrow{n \to \infty} \infty
\]

i.e. $X_n$ does not converge to $X$ in quadratic mean.
b) Let $Y_n = X_n - X$. Then $P(Y_n = 0|X) = 1 - 1/n$ and $P(Y_n = e^n - X|X) = 1/n$. Thus, the unconditional distribution of $Y_n$ is defined by $P(Y_n = 0) = 1 - 1/n$, $P(Y_n = e^n - 1) = P(Y_n = e^n + 1) = 1/(2n)$. Therefore

$$\mathbb{E}[(X_n - X)^2] = \mathbb{E}[Y_n^2] = \frac{1}{2n} (e^n - 1)^2 + \frac{1}{2n} (e^n + 1)^2 \to \infty$$

i.e. $X_n$ does not converge to $X$ in quadratic mean.

- **Convergence in probability**
  Notice that, $\forall \varepsilon > 0$, $|X_n - X| > \varepsilon \iff X_n \neq X$ so that $P(|X_n - X| > \varepsilon) = 1 - P(X_n \neq X) = \frac{1}{n} \to 0$ and hence $X_n \xrightarrow{p} X$.

- **Convergence in distribution**
  Since $X_n \xrightarrow{d} X$, then $X_n \xrightarrow{d} X$.

**Problem 8: Wasserman 5.14 [0 points]**

By the Central Limit Theorem

$$\sqrt{n}(\bar{X}_n - 1/2) \xrightarrow{d} N(0, 1/12)$$

where $\mathbb{E}[X_i] = 1/2$ and $V(X_i) = 1/12$. Now we can use the Delta method since the function $g(t) = t^2$ has first derivative $g'(t) = 2t$ with $g'(1/2) = 1 \neq 0$. Therefore

$$\sqrt{n}(Y_n - 1/4) \xrightarrow{d} N(0, 1/12)$$

**Remark**

Note that $\bar{X}_n$ and $Y_n = \bar{X}_n^2$ have the same asymptotic variance.