1. (a) Give an example of a sequence of random variables $X_n$, $X$ (all defined on a common probability space $(\Omega, \mathcal{A}, P)$) satisfying $X_n \to_{a.s.} X$, but $E(X_n) \not\to E(X)$.

(b) Give an example of a sequence of non-negative random variables $X_n$, $X$ on a common probability space satisfying $E(X_n) \to E(X)$ but $X_n \not\to_{a.s.} X$.

(c) Give an example of a sequence of random variables $X_n$, $X$ satisfying $X_n \to_{d} X$, but $X_n \not\to_{p,a.s.,1} X$.

Solution: (a) Let $U \sim \text{Uniform}[0, 1]$. For $\alpha \geq 0$, let $X_n \equiv n^\alpha 1_{(\frac{1}{n+1}, \frac{1}{n}]}(U)$.

Then $X_n \to_{a.s.} 0 \equiv X$ since $X_n = 0$ for $n > \frac{1}{U}$ and $P(U \in (0, 1]) = 1$.

But

$$E(X_n) = n^\alpha (n^{-1} - (n+1)^{-1}) = n^\alpha / (n+1) \to 1_{\{2\}}(\alpha) + \infty \cdot 1_{(2, \infty)}(\alpha).$$

Thus $E(X_n) \not\to E(X)$ for $\alpha \geq 2$.

(b) Let $U \sim \text{Uniform}[0, 1]$. For $0 < \alpha < 1$, $m \geq 1$, and $1 \leq k \leq 2^m$ define

$$Y_{m,k} \equiv (2^m)^{\alpha} 1_{(\frac{k-1}{2^m}, \frac{k}{2^m})}(U);$$

Then let $X_n \equiv X_{2^m + k} \equiv Y_{m,k}$ for $m \geq 1$ and $1 \leq k \leq 2^m$. Note that

$$E(X_n) = E(Y_{m,k}) = 2^m \cdot 2^{-m} = 2^{(\alpha-1)m} \to 0 \text{ as } n = 2^m + k \to \infty,$$

but $X_n = Y_{m,k} > 0$ i.o. with probability 1 since $P(U \in (0, 1]) = 1$.

(c) Suppose that $X_1, X_2, \ldots$ are independent and identically distributed random variables with common distribution function $F$ all defined on a common probability space. (We will make this completely rigorous in chapter 5.) Then $F_n(x) = P(X_n \leq x) = F(x)$, so $F_n$ certainly converges to $F$ for all $x \in \mathbb{R}$ and in particular at all $x \in C_F$. Thus $X_n \to_{d} X$, but $X_n \not\to_{a.s.,p,1} X$.

Alternatively, the $X_n$’s could be taken to be defined on separate probability spaces $(\Omega_n, \mathcal{A}_n, P_n)$ with induced distributions $P_{X_n}$ on $\mathbb{R}$ with distribution functions $F_n(x) \equiv P(X_n \leq x)$ satisfying $F_n \to_{d} F$. Now we cannot even talk about the random variables $X_n - X$, so $X_n \not\to_{a.s.,p,1} X$. Here is yet a further simple example: Suppose that $X_n \equiv U \sim \text{Uniform}[0, 1]$ for all $n \geq 1$. Suppose
that \( X \equiv 1 - U \). Then \( X \sim \text{Uniform}[0,1] \) so \( X_n \overset{d}{\rightarrow} X \) for all \( n \) and hence \( X_n \rightarrow_d X \), while \( X_n \equiv U \neq_a \text{a.s.} 1 - U \equiv X \), so \( X_n \overset{a.s.}{\rightarrow} X \). Note that

\[
\{|X_n - X| \geq \epsilon\} = \{|U - (1 - U)| \geq \epsilon\} = \{|2U - 1| \geq \epsilon\} = \{(1 - \epsilon)/2 \leq U \leq (1 + \epsilon)/2\}^c
\]

has \( P(|X_n - X| \geq \epsilon) = 1 - \epsilon \) for every \( n \geq 1 \) and hence \( X_n \overset{p}{\rightarrow} X \).

This also implies that \( X_n \overset{a.s.}{\rightarrow} X \).

2. PfS, Exercise 3.5.7, page 61, modified as follows: Suppose that \( f_0, f_1, \ldots \) are \( \geq 0 \), defined on a sigma-finite measure space \((\Omega, \mathcal{A}, \mu)\). (a) Suppose that \( \int_\Omega f_n d\mu = 1 \) for \( n = 0,1, \ldots \), and \( f_n \rightarrow_{a.e.} f_0 \) with respect to \( \mu \). Show that

\[
\sup_{A \in \mathcal{A}} \left| \int_A f_n d\mu - \int_A f_0 d\mu \right| \to 0 \quad \text{as} \quad n \to \infty.
\]

(b) Show that the conclusion of (a) holds if just \( f_n \rightarrow_{\mu} f_0 \) and \( \int_\Omega f_n d\mu \rightarrow \int_\Omega f_0 d\mu \).

**Solution:** (a) By the solution to problem \#3 below,

\[
\sup_{A \in \mathcal{A}} \left| \int_A f_n d\mu - \int_A f_0 d\mu \right| = \int (f_0 - f_n)^+ d\mu
\]

where \((f_0 - f_n)^+ \rightarrow_{a.e.} 0\) and is dominated by the integrable function \(f_0\). Hence the right side converges to 0 by the dominated convergence theorem.
(b) If we have $f_n \to \mu f_0$ and $\int f_n d\mu \to \int f_0 d\mu$, then we still have
\[
\sup_{A \in \mathcal{A}} |\int_A f_n d\mu - \int_A f_0 d\mu| \leq \sup_A \int_A |f_n - f_0| d\mu
\]
(1)

\[
\leq \int_\Omega |f_n - f_0| d\mu = \int_\Omega |f_0 - f_n| d\mu
\]
\[
= \int (f_0 - f_n)^+ d\mu + \int (f_0 - f_n)^- d\mu
\]
\[
= \int (f_0 - f_n)^+ d\mu + \int (f_0 - f_n)^+ d\mu - D_n
\]
\[
= 2 \int (f_0 - f_n)^+ d\mu - D_n
\]
(2)

where
\[
D_n \equiv \int_\Omega (f_0 - f_n) d\mu = \int_\Omega \{(f_0 - f_n)^+ - (f_0 - f_n)^-\} d\mu \to 0.
\]

But the right side of (2) converges to 0 by the dominated convergence theorem together with $D_n \to 0$.

3. Suppose that $P, Q$ are two probability measures on the same measurable space $(\Omega, \mathcal{A})$ which are both absolutely continuous with respect to the measure $\mu$ with densities (Radon-Nikodym derivatives) $p$ and $q$ respectively. Thus $P(A) = \int_A p d\mu$ and $Q(A) = \int_A q d\mu$ for $A \in \mathcal{A}$. Show that
\[
d_{TV}(P, Q) \equiv \sup_{A \in \mathcal{A}} |P(A) - Q(A)| = \frac{1}{2} \int |p - q| d\mu = \int (p - q)^+ d\mu.
\]

**Solution:** Let $\delta = p - q$, so that $\int_\Omega \delta d\mu = 0$. Then for $A \in \mathcal{A}$ we have
\[
0 = \int_\Omega \delta d\mu = \int_A \delta d\mu + \int_{A^c} \delta d\mu
\]
and hence $|\int_{A^c} \delta d\mu| = |\int_A \delta d\mu|$. Thus for $A \in \mathcal{A}$ we have
\[
2|\int_A \delta d\mu| = |\int_A \delta d\mu| + |\int_{A^c} \delta d\mu| \leq \int_\Omega |\delta| d\mu.
\]
If $A = [\delta \geq 0]$, then we have equality in the above inequality, and hence it follows that
\[
\sup_{A \in \mathcal{A}} |P(A) - Q(A)| = \sup_{A \in \mathcal{A}} \left| \int_A (p-q) d\mu \right| = \frac{1}{2} \int_{\Omega} |p-q| d\mu = \int (p-q)^+ d\mu.
\]
Note that the hypothesis of this problem, namely $P \ll \mu$ and $Q \ll \mu$ for some measure $\mu$ is always satisfied with $\mu \equiv P + Q$.

4. Suppose that $X_n \sim \text{Binomial}(n, p_n)$ for $n = 1, 2, \ldots$ with $np_n \to \lambda > 0$, and let $P_n$ be the induced distribution of $X_n$ on $\mathbb{R}$. Let $X_0 \sim \text{Poisson}(\lambda)$ and let $P_0$ be the corresponding induced distribution on $\mathbb{R}$. Use Scheffé’s theorem to show that $d_{TV}(P_n, P_0) \to 0$ as $n \to \infty$.

**Solution:** As we showed in problem 5, Problem Set 1,
\[
p_n(k) \equiv P(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda} \equiv p_0(k)
\]
for each fixed $k \geq 0$. Thus the hypotheses of problem 2(a) hold and we conclude that
\[
d_{TV}(P_n, P_0) = \sup_{A \in \mathcal{B}} |P_n(A) - P_0(A)| \to 0.
\]
This is considerably stronger than $X_n \to_d X_0 \sim \text{Poisson}(\lambda)$.

5. Let $X_{n1}, \ldots, X_{nn}$ be independent, $X_{nk} \sim \text{Bernoulli}(p_{nk})$, and let $Y_n \sim \text{Poisson} \left( \sum_{k=1}^{n} p_{nk} \right)$. Let $P_n$ be the distribution of $\sum_{k=1}^{n} X_{nk}$ and let $Q_n$ be the distribution of $Y_n$. Show that
\[
d_{TV}(P_n, Q_n) = \sup_{A \in \mathcal{B}} |P(S_n \in A) - P(Y_n \in A)| \leq \sum_{k=1}^{n} p_{nk}^2.
\]
Note that when $p_{nk} = p_n \to 0$ for all $k$ and $np_n \to \lambda$, then $\sum_{k=1}^{n} p_{nk}^2 = (np_n)^2/n = O(n^{-1})$.

**Hint:** Construct $S_n$ and $Y_n$ on a common probability space as follows: let $T_{nk} \sim \text{Poisson}(p_{nk})$, $k = 1, \ldots, n$ be independent, and let $Z_{nk} \sim \text{Bernoulli}(1 - (1 - p_{nk}) e^{p_{nk}})$, $k = 1, \ldots, n$ be independent and independent of the $T_{nk}$’s. Define $X_{nk} = 1_{[T_{nk} \geq 1]} + 1_{[T_{nk} = 0]} 1_{[Z_{nk} = 1]}$. Set
\[ S_n = \sum_{k=1}^n X_{nk}, \quad Y_n = \sum_{k=1}^n T_{nk}. \] Check that \( X_{nk} \sim \text{Bernoulli}(p_{nk}) \), \( Y_n \sim \text{Poisson}(\sum_{k=1}^n p_{nk}) \), and

\[
P(T_{nk} = 0, X_{nk} = 1) = e^{-p_{nk}} - (1 - p_{nk})
\]
\[
P(T_{nk} \geq 1, X_{nk} = 0) = 0, \quad P(T_{nk} \geq 2) = 1 - e^{-p_{nk}} - p_{nk} e^{-p_{nk}}.
\]

Show that

\[
d_{TV}(P_n, Q_n) \leq P(S_n \neq Y_n) \leq \sum_{k=1}^n P(X_{nk} \neq T_{nk}) \leq \sum_{k=1}^n p_{nk}^2.
\]

**Solution:** Using the notation in the hint we first show that \( X_{nk} \sim \text{Bern}(p_{nk}) \): this follows since, using \( P(Y_{\lambda} \geq 1) = 1 - P(Y_{\lambda} = 0) = 1 - e^{-\lambda} \) if \( Y_{\lambda} \sim \text{Poisson}(\lambda) \),

\[
P(X_{nk} = 1) = P(T_{nk} \geq 1) + P(T_{nk} = 0) P(Z_{nk} = 1)
\]
\[
= 1 - e^{-p_{nk}} + e^{-p_{nk}} (1 - (1 - p_{nk}) e^{p_{nk}})
\]
\[
= p_{nk}.
\]

Next,

\[
P(T_{nk} = 0, X_{nk} = 1) = e^{-p_{nk}} (1 - (1 - p_{nk}) e^{p_{nk}}) = e^{-p_{nk}} - (1 - p_{nk}),
\]

while

\[
P(T_{nk} \geq 1, X_{nk} = 0) = P(T_{nk} \geq 1, T_{nk} = 0) = 0,
\]

and

\[
P(T_{nk} \geq 2) = 1 - P(T_{nk} = 0 \text{ or } 1) = 1 - e^{-p_{nk}} - p_{nk} e^{-p_{nk}}.
\]

Thus

\[
P(X_{nk} \neq T_{nk}) = P(X_{nk} = 0, T_{nk} = 1) + P(X_{nk} = 1, T_{nk} = 0) + P(T_{nk} \geq 2)
\]
\[
= 0 + e^{-p_{nk}} - (1 - p_{nk}) + 1 - e^{-p_{nk}} - p_{nk} e^{-p_{nk}}
\]
\[
= p_{nk} (1 - e^{-p_{nk}}) \leq p_{nk}^2.
\]
Thus for any $A \in 2^\Omega$,

$$
P(S_n \in A) - P(Y_n \in A) = P([S_n \in A] \cap [S_n = Y_n]) + P([S_n \in A] \cap [S_n \neq Y_n]) - P([Y_n \in A] \cap [S_n = Y_n]) - P([Y_n \in A] \cap [S_n \neq Y_n])
$$

$$
\leq P([S_n \in A] \cap [S_n \neq Y_n]) = P(S_n \neq Y_n).
$$

Similarly, by a symmetric argument,

$$
P(S_n \in A) - P(Y_n \in A) \geq -P(S_n \neq Y_n),
$$

and this yields

$$
d_{TV}(P_n, Q_n) = \sup_{A \in 2^\Omega} |P(S_n \in A) - P(Y_n \in A)|
$$

$$
\leq P(S_n \neq Y_n) \leq \sum_{k=1}^n P(X_{nk} \neq T_{nk}) \leq \sum_{k=1}^n p_{nk}^2.
$$

If $p_{nk} = \lambda/n$ for $1 \leq k \leq n$ for some $\lambda > 0$, then this bound yields

$$
d_{TV}(P_n, Q_n) \leq n(\lambda/n)^2 = \frac{\lambda^2}{n}.
$$

For still stronger results, see Barbour, Holst, and Janson (1992), Poisson Approximation.

6. Let $X_1, X_2, \ldots, X_n, \ldots$ be i.i.d. Uniform$(0,1)$ random variables. Let $Y_n = nX_{1:n}$ be the first order statistic of the first $n$ of the $X_i$'s.

(a) Compute the survival function $1 - F_{Y_n}$ of $Y_n$ and show that $Y_n \to_d Y \sim \text{exponential}(1)$.

(b) Compute the density function $f_{Y_n}$ of $Y_n$ and show that $f_{Y_n}(y) \to f_Y(y) = e^{-y}$ for $y \geq 0$.

(c) Use (b) and Scheffé’s theorem to show that

$$
d_{TV}(P_{Y_n}, P_Y) = \frac{1}{2} \int_0^\infty |f_{Y_n}(y) - f_Y(y)|dy \to 0.
$$

How fast is the convergence in the last display?
**Solution:** (a) For \( y \geq 0 \) we have

\[
1 - F_{Y_n}(y) = P(Y_n > y) = P(X_{n:1} > y/n)
= P(X_1 > y/n, \ldots, Y_n > y/n) = P(X_1 > y/n) \cdots P(X_n > y/n)
= P(X_1 > y/n)^n = (1 - y/n)_+
\]

where \( z_+ = z \{ z \geq 0 \} \).

(b) It follows from the last line of (a) that

\[
1 - F_{Y_n}(y) = (1 - y/n)_+^n \to e^{-y} = P(Y > y) \text{ for all } y \geq 0
\]

where \( Y \sim \text{Exponential}(1) \). Thus \( Y_n \to_d Y \).

(b) The density function of \( Y_n \) is just

\[
f_{Y_n}(y) = F'_{Y_n}(y) = (1 - y/n)_+^{n-1}1_{[0,\infty)}(y)
\to e^{-y} = f_Y(y) \text{ for all } y > 0.
\]

(c) The hypotheses of Scheffé’s theorem are satisfied by (b), so we conclude that

\[
d_{TV}(P_{Y_n}, P_Y) = \frac{1}{2} \int_0^\infty |f_{Y_n}(y) - f_Y(y)|dy \to 0.
\]

To start to get a handle on how fast the right side in the last display converges to 0, first note that

\[
n(f_{Y_n}(y) - f_Y(y)) \to \frac{1}{2}y(2 - y)e^{-y} \equiv h(y)
\]

for each fixed \( y \in (0, \infty) \) where \( \int_0^\infty |h(y)|dy = 4e^{-2} \). Thus if we can justify the interchange of limit and integration it would follow that

\[
n d_{TV}(P_{Y_n}, P_Y) = \frac{n}{2} \int_0^\infty |f_{Y_n}(y) - f_Y(y)|dy
\to \frac{1}{2} \int_0^\infty \frac{1}{2}y(2 - y)e^{-y}dy
= 2e^{-2}.
\]

I suspect that this limiting relation can be turned into an exact bound with a little more work.
7. Suppose that \( g : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\} \) is convex. Show that \( h(x,t) \equiv tg(x/t) \) is a convex function on \( \mathbb{R}^d \times (0, \infty) \).

**Hint:** First show that \( h(cx, st) = ch(x, t) \) for any \( c > 0, \ (x, t) \in \mathbb{R}^d \times (0, \infty) \), and hence \( t^{-1}h(x, t) = h(x/t, 1) \).

**Solution:** First note that if \( c > 0 \), then \( h(cx, ct) = ctg(cx/ct) = ch(x, t) \), so with \( c = 1/t \) it follows that \( g(x/t) = h(x/t, 1) = t^{-1}h(x, t) \).

Now let \( \lambda \in [0, 1] \) and \( (x_j, t_j) \in \mathbb{R}^d \times (0, \infty) \) for \( j = 1, 2 \). Then

\[
\begin{align*}
    h(\lambda x_1 + \lambda x_2, \lambda t_1 + \lambda t_2) & = (\lambda t_1 + \lambda t_2)g \left( \frac{\lambda x_1 + \lambda x_2}{\lambda t_1 + \lambda t_2} \right) \\
    & = (\lambda t_1 + \lambda t_2)g \left( \frac{\lambda t_1 (x_1/t_1) + \lambda t_2 (x_2/t_2)}{\lambda t_1 + \lambda t_2} \right) \\
    & \leq (\lambda t_1 + \lambda t_2) \left\{ \frac{\lambda t_1}{\lambda t_1 + \lambda t_2}g(x_1/t_1) + \frac{\lambda t_2}{\lambda t_1 + \lambda t_2}g(x_2/t_2) \right\} \\
    & = \lambda t_1 g(x_1/t_1) + \lambda t_2 g(x_2/t_2) = \lambda h(x_1,t_1) + \lambda h(x_2,t_2).
\end{align*}
\]

It follows that \( h \) is convex.

8. For \( s \in \mathbb{R} \cup \{\pm \infty\} \), \( u, v \in \mathbb{R}^+ \), and \( \lambda \in [0, 1] \), the H"older mean (or generalized mean) \( M_s(u, v; \lambda) \) of order \( s \) is defined by

\[
M_s(u, v; \lambda) = \begin{cases} 
(\lambda u^s + (1 - \lambda)v^s)^{1/s}, & s \neq 0, \ u, v > 0, \\
0, & s < 0, \ uv = 0, \\
u^\lambda v^{1-\lambda}, & s = 0, \\
u \wedge v, & s = -\infty, \\
u \vee v, & s = +\infty.
\end{cases}
\]

(a) Interpret \( M_s(u, v; \lambda) \) in terms of some function of the expected value of some random variable \( X \).

(b) Show that for any \( r < s \) the inequality \( M_r(u, v; \lambda) \leq M_s(u, v; \lambda) \) holds for all \( u, v \in \mathbb{R}, \ \lambda \in [0, 1] \). Thus

\[
M_r(u, v; \lambda) \leq M_0(u, v; \lambda) \leq M_s(u, v; \lambda).
\]

(In class on 10/31 we proved a related statement with \( r = -1 \) and \( s = 1 \).)
**Solution:** (a) Let $X$ be a random variable taking on the value $u$ with probability $\lambda$ and $v$ with probability $1 - \lambda$. Thus for $u, v > 0$ and $s \neq 0$ we have

$$M_s(u, v; \lambda) = \{E(X^s)\}^{1/s} \equiv \|X\|_s. \quad (3)$$

For $s = 0$ and $u, v > 0$

$$M_0(u, v; \lambda) = \exp(E \log X).$$

(b) For $0 < r < s$ the inequality $M_r(u, v; \lambda) \leq M_s(u, v; \lambda)$ becomes $\|X\|_r \leq \|X\|_s$ in view of (3), and this is just Liapunov’s inequality Inequality 3.4.4, PfS page 48.) For $r < s < 0$, the inequality follows by replacing $X$ by $1/X$ and $r < s < 0$ by $0 < -s < -r$. Now consider $r = 0$ and $0 < s$. By concavity of $g(y) = s^{-1} \log y$ it follows from Jensen’s inequality that

$$E \log X = Es^{-1} \log X^s \leq s^{-1} \log E(X^s) = \log(\|X\|_s),$$

and hence

$$M_0(u, v; \lambda) = E \exp(E \log X) \leq \|X\|_s = M_s(u, v; \lambda)$$

For the case $r = 0 = s$, note that $g(y) = r^{-1} \log y$ is convex, so Jensen’s inequality gives

$$E \log X = E(r^{-1} \log X^r) \geq r^{-1} \log X^r = \log(\|X\|_r),$$

and hence

$$M_0(u, v; \lambda) = E \exp(E \log X) \geq \|X\|_r = M_r(u, v; \lambda).$$

The cases with $r = -\infty$ or $s = \infty$ are easy. When $r < \infty$ and $s = +\infty$,

$$\{EX^r\}^{1/r} \leq \{EX^t\}^{1/t} = \{\lambda u^t + \lambda v^t\}^{1/t} \quad \text{for every} \quad r < t < \infty$$

$$\leq \{\lambda(u \lor v)^t + \lambda(u \lor v)^t\}^{1/t} = u \lor v = M_\infty(u, v; \lambda).$$

The case $r = -\infty$ and $s > -\infty$ is similar.
9. PfS, Exercise 4.1.2, page 67: Identify $\phi^+$, $\phi^-$, $|\phi|$ and $|\phi|(\Omega)$ in the context of the prototypical situation of example 4.1.1, page 66. Be sure to specify $\Omega^+$ and $\Omega^-$.  

**Solution:** I claim that 

$$
\phi^+(A) = \int_A X^+ d\mu = \phi(A\Omega^+) \text{ with } \Omega^+ = \{\omega : X(\omega) \geq 0\},
$$

$$
\phi^-(A) = \int_A X^- d\mu = -\phi(A\Omega^-) \text{ with } \Omega^- = \{\omega : X(\omega) < 0\},
$$

$$
|\phi|(A) = \int_A |X|d\mu, \quad \text{and}
$$

$$
|\phi|(\Omega) = \int |X|d\mu.
$$

To see this, note that $\Omega^+$, $\Omega^-$ are, respectively, positivity, negativity sets for $\phi$ since

$$
\phi(A) = \int_A X d\mu \geq 0 \quad \text{for all events } A \subset \Omega^+,
$$

$$
\phi(A) = \int_A X d\mu \leq 0 \quad \text{for all events } A \subset \Omega^-.
$$

Furthermore, if $\tilde{\Omega}^+$, $\tilde{\Omega}^-$ denote the decomposition guaranteed by the Jordan-Hahn theorem 1.1, then

$$
\phi(\Omega^+ \setminus \tilde{\Omega}^+) = \phi(\Omega^+ \cap \tilde{\Omega}^-) = 0, \quad \text{and}
$$

$$
\phi(\tilde{\Omega}^+ \setminus \Omega^+) = \phi(\tilde{\Omega}^+ \cap \Omega^-) = 0,
$$

where the zeroes follow by using the definitions of $\Omega^+$, $\Omega^-$, $\tilde{\Omega}^+$, $\tilde{\Omega}^-$. Thus

$$
|\phi|(\Omega^+ \Delta \tilde{\Omega}^+) = 0;
$$

i.e. $\Omega^+ = [X \geq 0]$ differs from $\tilde{\Omega}^+$ by (at most) a set of $|\phi|$—measure 0.