Statistics 522, Midterm Exam Solutions

Wellner; 2/19/2020

1. (40 points). **Define** *four* of the following six terms:
   (a) The *tail sigma-field* of an arbitrary sequence of random variables $X_1, X_2, X_3, \ldots$.
   (b) *Independent sub-sigma-fields $\mathcal{A}_1, \ldots \mathcal{A}_n$ of a sigma-field $\mathcal{A}$* where $(\Omega, \mathcal{A}, P)$ is a fixed probability space.
   (c) The conditional expectation of a random variable X given a sub-sigma-field $\mathcal{D}$.
   (d) A martingale, sub-martingale, and super-martingale.
   (e) A stopping time T (relative to a filtration $\mathcal{A}_n$).
   (f) A Brownian motion process $\mathbb{S}$ on $[0, \infty)$.

   Solution: See course notes.

2. (40 points). Give careful **statements** of *four* of the following eight theorems or results:
   (a) Kolmogorov's zero-one law.
   (b) The *two-series theorem*.
   (c) The *conditional version of Jensen's inequality*.
   (d) The *first Borel-Cantelli lemma*.
   (e) The *second Borel-Cantelli lemma*.
   (f) Doob's decomposition theorem for sub-martinagles.
   (g) The step-wise smoothing property of conditional expectations.
   (h) Kolmogorov's maximal inequality for sums of independent random variables with finite second moments.

   Solution: See course notes.

3. (40 points). Suppose that $X \in L_2(\Omega, \mathcal{A}, P)$ and $\mathcal{D}$ is a sub-sigma field of $\mathcal{A}$. The conditional variance of X given $\mathcal{D}$ is defined by

$$Var(X|\mathcal{D}) = E\left\{(X - E(X|\mathcal{D}))^2|\mathcal{D}\right\} .$$

   (a) Prove that $Var(X) = E[Var(X|\mathcal{D})] + Var(E(X|\mathcal{D}))$.
   (b) Show that $E(X - Z)^2$ is minimized over all $\mathcal{D}-$measurable random variables Z by $E(X|\mathcal{D})$.
   (c) Interpret the formula in (a) geometrically.
   (d) Let $Y_1, Y_2, \ldots$ be i.i.d. random variables with $E(Y_i) = \mu$ and $Var(Y_i) = \sigma^2$. Let N be independent of the Y_i's and consider $S_N \equiv Y_1 + \cdots + Y_N$. Show that: $E(S_N) = \mu E(N)$ and $Var(S_N) = \mu^2 Var(N) + \sigma^2 E(N)$.
   (e) Does the formula for $E(S_N)$ in (d) extend to a stopping time N with $E(N) < \infty$?

Solution: (a) By adding and subtracting $E(X|\mathcal{D})$ we get

$$\begin{aligned} Var(X) &= E(X - E(X|\mathcal{D}) + E(X|\mathcal{D}) - E(X))^2 \\ &= E(X - E(X|\mathcal{D}))^2 + E\{(E(X|\mathcal{D}) - E(X))^2\} \\ &\qquad + 2E\{(X - E(X|\mathcal{D}))(E(X|\mathcal{D}) - E(X))\} \\ &= E\{E[(X - E(X|\mathcal{D}))^2|\mathcal{D}]\} + Var[E(X|\mathcal{D})] \\ &= E\{Var[X|\mathcal{D}]\} + Var[E(X|\mathcal{D})] \end{aligned}$$

since

$$\begin{aligned} & E\{(X - E(X|\mathcal{D}))(E(X|\mathcal{D}) - E(X))\} \\ &= E\{E[(X - E(X|\mathcal{D}))(E(X|\mathcal{D}) - E(X))|\mathcal{D}]\} \\ &= E\{(E(X|\mathcal{D}) - E(X))E[X - E(X|\mathcal{D})|\mathcal{D}]\} \\ &= E\{(E(X|\mathcal{D}) - E(X))(E[X|\mathcal{D}] - E[X|\mathcal{D}])\} \\ &= E\{(E(X|\mathcal{D}) - E(X)) \cdot 0\} = 0\,. \end{aligned}$$

(b) To show that $E(X - Z)^2$ is minimized over all $\mathcal{D}$-measurable functions Z by $E(X|\mathcal{D})$, write

$$\begin{aligned} E(X - Z)^2 &= E(X - E(X|\mathcal{D}) + E(X|\mathcal{D}) - Z)^2 \\ &= E[(X - E(X|\mathcal{D}))^2] + E[(E(X|\mathcal{D}) - Z)^2] \\ &\qquad + 2E[(X - E(X|\mathcal{D}))(E(X|\mathcal{D}) - Z)] \\ &= E[(X - E(X|\mathcal{D}))^2] + E[(E(X|\mathcal{D}) - Z)^2] \\ &\geq E[(X - E(X|\mathcal{D}))^2] \end{aligned}$$

with equality in the last inequality if and only if $Z = E(X|\mathcal{D})$. Here the third equality holds since

$$\begin{aligned} & E[(X - E(X|\mathcal{D}))(E(X|\mathcal{D}) - Z)] \\ &= E\{E[(X - E(X|\mathcal{D}))(E(X|\mathcal{D}) - Z)|\mathcal{D}]\} \\ &= E\{(E(X|\mathcal{D}) - Z)E[X - E(X|\mathcal{D})|\mathcal{D}]\} \\ &= E\{(E(X|\mathcal{D}) - Z)(E[X|\mathcal{D}] - E(X|\mathcal{D}))\} \\ &= E\{(E(X|\mathcal{D}) - Z) \cdot 0\} = 0\,. \end{aligned} \tag{1}$$

(c) When $E(X) = 0$ the identity of (a) becomes

$$E(X^2) = E[(X - E(X|\mathcal{D}))^2] + E\{E[X|\mathcal{D}]^2\}\,.$$

This can be seen geometrically as a "Pythagorean" relationship as follows: the squared length of X, namely $E(X^2)$, has been decomposed into the sum of the squared length of t he $\mathcal{D}-$ measurable predictor $E(X|\mathcal{D})$ and the squared length of the prediction error

$X - E(X|\mathcal{D})$. Furthermore, by (1), the prediction error $X - E(X|\mathcal{D})$ is orthogonal to all the $\mathcal{D}-$measurable predictors Z:

$$\begin{aligned} E\{(X - E(X|\mathcal{D}))Z\} &= E\{Z(X - E(X|\mathcal{D}))\} \\ &= E\{E[Z(X - E(X|\mathcal{D}))|\mathcal{D}]\} \\ &= E\{ZE[X - E(X|\mathcal{D})|\mathcal{D}]\} \\ &= E\{Z(E(X|\mathcal{D}) - E(X|\mathcal{D}))\} = E\{Z \cdot 0\} = 0. \end{aligned}$$

(d) Here $E(S_N|N) = E(\sum_{i=1}^N X_i) = N \cdot \mu$ and $Var(S_N|N) = N\sigma^2$ and hence by (a)

$$\begin{aligned} Var(S_N) &= E(Var(S_N|N)) + Var(E(S_N|N)) \\ &= E(N\sigma^2) + Var(N\mu) = E(N)\sigma^2 + Var(N)\mu^2. \end{aligned}$$

(e) Yes. As we saw in class on 12 February, if the X_i's are independent and identically distributed and N is a stopping time relative to $\mathcal{A}_n \equiv \sigma[X_1, \ldots, X_n]$, by writing $S_N = \sum_{k=1}^\infty X_k 1\{k \le N\}$ where $\{N \ge k\} = \{N \le k-1\}^c$ we see that X_k and $1\{k \le N\}$ are independent. Thus

$$\begin{aligned} E(S_N) &= E\left(\sum_{k=1}^\infty X_k 1\{k \le N\}\right) \\ &= \sum_{k=1}^\infty E(X_k 1\{k \le N\} = \sum_{k=1}^\infty E(X_k)E(1\{k \le N\}) \\ &= \mu \sum_{k=1}^\infty P(N \ge k) = \mu E(N). \end{aligned}$$

To justify the interchange of expectation and summation, apply the same argument with X_k replaced by $|X_k|$.

Do one of the following two problems: 4 or 5

4. (40 points). Suppose that $\xi_1, \xi_2, \ldots$ are i.i.d. Uniform$(0,1)$ random variables, and let $\mathbb{G}_n(t) = n^{-1}\sum_{i=1}^n 1_{[0,t]}(\xi_i)$ be the empirical distribution function of the ξ_i's. Define $\mathcal{F}_n(s) \equiv \sigma\{\mathbb{G}_n(t) : \ s \le t \le 1\}$ for $0 < s \le 1$. (Note the correction to the original problem statement here.
(a) Show that $\{\mathbb{G}_n(t)/t, \mathcal{F}_n(t) : \ 0 < t \le 1\}$ is a reverse martingale in the following sense: if $0 < s < t < 1$, then $E\left\{\mathbb{G}_n(s)/s \middle| \mathcal{F}_n(t)\right\} = \mathbb{G}_n(t)/t$.
Hint: Note that $(n\mathbb{G}_n(s), n(\mathbb{G}_n(t) - \mathbb{G}_n(s)), n(1 - \mathbb{G}_n(t)))$ has a multinomial distribution $\text{Mult}_3(n, (s, t-s, 1-t))$, and hence the
conditional distribution of $(n\mathbb{G}_n(s), n(\mathbb{G}_n(t) - \mathbb{G}_n(s)))$ given $\mathcal{F}_n(t)$ is the same as the conditional distribution of $(n\mathbb{G}_n(s), n(\mathbb{G}_n(t) - \mathbb{G}_n(s)))$ given $n(1 - \mathbb{G}_n(t))$, namely $\text{Mult}_2(n\mathbb{G}_n(t), (s/t, 1 - s/t))$.

(b) Use the result of (a) to show that for $0 < a < 1$ and $\lambda > 0$

$$P(\sup_{a\le t\le 1} \frac{\mathbb{G}_n(t)}{t} > \lambda) \le \exp(-nah(\lambda))\,.$$

where $h(\lambda) = \lambda(\log\lambda - 1) + 1$.
(c) Use the result of (b) to show that if $a = a_n$ satisfies $na_n \to \infty$, then

$$\sup_{a_n\le t\le 1} \frac{\mathbb{G}_n(t)}{t} \to_p 1\,.$$

[Hint: take $\lambda = 1+\epsilon$ with $\epsilon > 0$ in (b) and note that $h(1+\epsilon) > 0$.]

Solution: (a) By the hint, the conditional distribution of $n\mathbb{G}_n(s)$ given $n\mathbb{G}_n(t)$ is Binomial $(n\mathbb{G}_n(t), s/t)$, and thus

$$E\left\{n\mathbb{G}_n(s)\Big|\mathcal{F}_n(t)\right\} \;=\; E\left\{n\mathbb{G}_n(s)\Big|n\mathbb{G}_n(t)\right\} = n\mathbb{G}_n(t)\cdot\frac{s}{t} \quad \text{a.s.;}$$

dividing across this identity by $n\cdot s$ yields the claim:

$$E\left\{\frac{\mathbb{G}_n(s)}{s}\Big|\mathcal{F}_n(t)\right\} = \frac{\mathbb{G}_n(t)}{t} \quad \text{a.s..}$$

(b) Since $\{\mathbb{G}_n(t)/t, \mathcal{F}_n(t)\}_{0<t\le 1}$ is a reverse martingale and $\varphi_r(x) \equiv e^{rx}$ is a convex function of x for each $r \in \mathbb{R}$, it follows that $\{\varphi_r(\mathbb{G}_n(t)/t), \mathcal{F}_n(t)\}_{0<t\le 1}$ is a reverse sub-martingale. Thus by Doob's maximal inequality it follows that for $r > 0$ we have

$$\begin{aligned}
P\left(\sup_{a\le t\le 1}\frac{\mathbb{G}_n(t)}{t} > \lambda\right) &= P\left(\sup_{a\le t\le 1}\exp\left(r\mathbb{G}_n(t)/t\right) > e^{r\lambda}\right)\\
&\le \frac{E\exp(r\mathbb{G}_n(a)/a)}{\exp(r\lambda)}.
\end{aligned}$$

Now the expectation in the numerator on the right side is just

$$\begin{aligned}
E\exp\left((r/na)n\mathbb{G}_n(a)\right) &= \left(E\exp((r/na)1_{[0,a]}(\xi_1))\right)^n\\
&= \left((1-a)\exp((r/na)\cdot 0) + a\exp((r/na))\right)^n\\
&= \left((1-a) + ae^{r/(na)}\right)^n\\
&= \left(1 - a(1-e^{r/(na)})\right)^n\\
&\le \exp\left(-na(1-e^{r/(na)})\right)
\end{aligned}$$

Combining this with the denominator gives

$$\begin{aligned}
P\left(\sup_{a\le t\le 1}\frac{\mathbb{G}_n(t)}{t} > \lambda\right) &\le \exp\left(-na(1-e^{r/(na)}) - r\lambda\right)\\
&= \exp\left(-na(1-e^{r/(na)} + (r/na)\lambda)\right).
\end{aligned}$$

for each $r > 0$. Minimizing this with respect to $r > 0$ we find that the minimizing value of r is $r^* \equiv na \log \lambda$. The resulting bound is found to be

$$P\left(\sup_{a \le t \le 1} \frac{\mathbb{G}_n(t)}{t} > \lambda \right) \le \exp(-nah(\lambda))$$

(c) The function h is strictly convex with $h(1) = 0$, $h(0) = 1$, $h'(\lambda) = \log(\lambda)$ and $h''(\lambda) = 1/\lambda > 0$, and therefore $h(1+\epsilon) > 0$ for each $\epsilon > 0$. Thus for each fixed $\epsilon > 0$ we find that

$$P\left(\sup_{a_n \le t \le 1} \frac{\mathbb{G}_n(t)}{t} > 1 + \epsilon \right) \le \exp(-na_n h(1+\epsilon)) \to 0$$

if $na_n \to \infty$. Since $\sup_{a_n \le t \le 1}(\mathbb{G}_n(t)/t) \ge \mathbb{G}_n(1)/1 = 1$, it follows that $\sup_{a_n \le t \le 1}(\mathbb{G}_n(t)/t) \to_p 1$.

5. (40 points) Let $p \in (0,1)$. Suppose that $\{Z_n\}_{n=0}^{\infty}$ is a sequence of random variables with

$$P(Z_{n+1} = 2j | Z_n = i) = \binom{i}{j} p^j (1-p)^{i-j}, \quad j \in \{0, \ldots, i\}, i \in \{0, 2, 4, \ldots\}. \tag{2}$$

with the convention that $P(Z_{n+1} = 0 | Z_n = 0) = 1$. Also assume that $P(Z_0 = k_0) = 1$ for a fixed (possibly large) even integer $k_0 \ge 2$. Let $\mathcal{A}_n \equiv \sigma[Z_0, Z_1, \ldots, Z_n]$, $n \ge 1$.
(a) Interpret the sequence $\{Z_n\}$ in terms of Binomial and Bernoulli random variables.
(b) Show that when $p = 1/2$ the process $\{Z_n, \mathcal{A}_n\}$ is a martingale with mean k_0 (with respect to the filtration $\{\mathcal{A}_n\}$ with $\mathcal{A}_n = \sigma\{Z_0, \ldots, Z_n\}$ for $n \ge 0$). For what values of p is $\{Z_n, \mathcal{A}_n\}$ a sub-martingale? For what values of p is it a super-martingale?
For the rest of the problem, assume that $p = 1/2$.
(c) Show that with $Y_n \equiv P(Z_{n+1} = 0 | \mathcal{A}_n) = P(Z_{n+1} = 0 | Z_n)$, the process $\{Y_n, \mathcal{A}_n\}_{n \ge 0}$ is a sub-martingale. In fact, use Jensen's inequality to show that $\{Y_n, \mathcal{A}_n\}_{n \ge 0}$ is an almost surely strictly increasing sub-martingale in the sense that $E(Y_{n+1} | \mathcal{A}_n) > Y_n$ almost surely.
Hint: Express Y_n in terms of an exponential function.
(d) Use the result of (d) to show that $Y_n \to_{a.s.} 1$ and hence that $Z_n \to_{a.s.} 0$.

Solution: (a) Note that (2) implies that

$$Z_{n+1} \stackrel{d}{=} 2\text{Binomial}(Z_n, p) \ \text{ conditionally on } \ Z_n.$$

Furthermore, a Binomial(m, p) is equal in distribution to $\sum_{j=1}^{m} B_j$ where the B_j's are i.i.d. Bernoulli(p) random variables.
(b) It follows easily from (a) that

$$E(Z_{n+1} | \mathcal{A}_n) = E(Z_{n+1} | Z_n) = 2 Z_n p \ \text{ a.s.}$$

Thus $\{Z_n, \mathcal{A}_n\}_{n\geq 0}$ is a martingale (with mean $k_0 = Z_0$) if $p = 1/2$, while it is a sub-martingale or super-martingale depending on whether $p > 1/2$ or $p < 1/2$.
(c) From the formula (2) it follows that $Y_n \equiv P(Z_{n+1} = 0|Z_n) = \binom{Z_n}{0} p^0 (1-p)^{Z_n} = \exp(-Z_n \log(1/(1-p)))$ where $\log(1/(1-p)) > 0$ for $p \in (0,1)$. When $p = 1/2$ this becomes $Y_n \equiv \exp(-Z_n \log 2)$. Since the function $\varphi(x) = \exp(-x \log 2)$ is convex and Z_n is a non-negative martingale it follows that Y_n is a sub-martingale; in fact, since φ is strictly convex and the Binomial distribution is not concentrated at its mean, $E(Y_{n+1}|\mathcal{A}_n) = E(\exp(-Z_{n+1}\log 2)|\mathcal{A}_n) > \exp(-Z_n \log 2) = (1/2)^{Z_n} = Y_n$ almost surely. Furthermore, note that

$$Es^{\text{Binomial}(n,p)} = (1-p+ps)^n.$$

Thus

$$\begin{aligned}
EY_{n+1} &= E(1/2)^{Z_{n+1}} = E\{E((1/2)^{Z_{n+1}}|Z_n)\} = E\{[(1/2)(1+(1/4))]^{Z_n}\} \\
&= E(s_1^{Z_n}) = E\{[(1/2)(1+s_1)]^{Z_{n-1}}\} \ \text{ with } \ s_1 \equiv (1/2)(1+1/4) \equiv f(1/2) \\
&\qquad \text{where } \ f(s) \equiv (1/2)(1+s^2) \\
&= Es_2^{Z_{n-1}} \ \text{ with } \ s_2 \equiv f(s_1) \\
&= \cdots = s_{n+1}^{Z_0} \quad \text{where } \ s_{n+1} \equiv f(s_n)
\end{aligned}$$

Thus $s_{n+1} \to s_\infty$ where $s_\infty = f(s_\infty) = (1/2)(1+s_\infty^2)$. Since the only fixed point of f is 1, we find that $s_\infty = 1$ and hence $E(Y_{n+1}) \to 1$.
(d) Thus $\{Y_n, \mathcal{A}_n\}_{n\geq 0}$ is an integrable sub-martingale, and by the s-martingale convergence theorem, $Y_n \to_{a.s.} Y_\infty$. It follows that $Y_\infty = 1$ and this implies that $Z_n \to_{a.s.} 0$.

Remark: Since $Z_{n+1} \stackrel{d}{=} \sum_{j=1}^{Z_n} X_{n,j}$ where the $X_{n,j}$'s are i.i.d. 2Bernoulli(1/2) with $m = E(X_{1,1}) = 1$, this is a special case of the branching process example 13.4.5, page 366, PfS with basic martingale having mean k_0 rather than 1. Note that (c) and (d) give a different proof of Theorem 13.4.1(ii) in this special case.