

Statistics 523, Midterm Exam Solutions

Wellner; 5/15/2017

1. (30 points). **Define** *three* of the following five terms:
 - (a) An infinitely divisible random variable X
(or distribution function F).
 - (b) A stable distribution F (or stable random variable) X .
 - (c) The domain of attraction of a stable distribution G .
 - (d) The *strong Markov property* of a process $\{X(t) : 0 \leq t < \infty\}$.
 - (e) A compound Poisson distribution.

Solution: See PfS and course notes.

2. (36 points). Give careful **statements** of *three* of the following five theorems or results:
 - (a) The Cramér-Lévy continuity theorem.
 - (b) Stein's identity characterizing the $N(0, 1)$ distribution.
 - (c) Four properties of Brownian motion \mathbb{S} on $[0, \infty)$.
 - (d) A result concerning embedding of one random variable with $E(X) = 0$ and variance $Var(X) = 1$ in Brownian motion.
 - (e) A result about embedding the partial sum process of $X_{n,1}, \dots, X_{n,n}$ i.i.d. F with $E(X_{n,i}) = 0$ and $Var(X_{n,i}) = 1$ in Brownian motion.

Solution: See PfS and course notes.

3. (24 points) Consider an urn with N balls labelled $\{a_1, \dots, a_N\}$ where $a_i \in \mathbb{R}$. Let X_1, \dots, X_n be a sample drawn without replacement from the urn.
- Describe the representation of X_1, \dots, X_n in terms of a random permutation $\underline{R} = (R_1, \dots, R_N)$ of the first N integers with $P(\underline{R} = \underline{r}) = 1/N!$ for each permutation $\underline{r} = (r_1, \dots, r_N)$ of $\{1, \dots, N\}$.
 - If $\bar{X}_n = n^{-1}(X_1 + \dots + X_n)$ is the sample mean of the X_i 's, compute $E(\bar{X}_n)$ in terms of the a_i 's.
 - What is $Var(\bar{X}_n)$ and how does it differ from the variance of the sample mean when the sampling is carried out *with replacement*?
 - Under what additional conditions would you expect to be able to prove that $(\bar{X}_n - E(\bar{X}_n))/\sqrt{Var(\bar{X}_n)} \rightarrow_d N(0, 1)$? (State a relevant theorem if you know one.)

Solution: (a) It is easily seen that

$$(X_1, \dots, X_n) \stackrel{d}{=} (a_{R_1}, \dots, a_{R_n})$$

where $\underline{R} = (R_1, \dots, R_N)$ is a random permutation of $\{1, \dots, N\}$.

(b) $E(\bar{X}_n) = n^{-1} \sum_{i=1}^n E a_{R_i} = E a_{R_1}$ by symmetry where $E a_{R_1} = N^{-1} \sum_{j=1}^N a_j \equiv \bar{a}_N$ is the population mean.

(c) It is easily calculated that

$$\begin{aligned} Var\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n Var(a_{R_i}) + \sum_{i \neq j} Cov(a_{R_i}, a_{R_j}) \\ &= n Var(a_{R_1}) + n(n-1) Cov(a_{R_1}, a_{R_2}) \end{aligned} \quad (1)$$

by symmetry where

$$Var(a_{R_1}) = N^{-1} \sum_{j=1}^N a_j^2 - \bar{a}_N^2 = N^{-1} \sum_{j=1}^N (a_j - \bar{a}_N)^2 \equiv \sigma_a^2$$

is the population variance. But taking $n = N$ in (1) yields

$$0 = Var\left(\sum_{i=1}^N a_i\right) = Var\left(\sum_{i=1}^N a_{R_i}\right) = N\sigma_a^2 + N(N-1)Cov(a_{R_1}, a_{R_2}),$$

and this yields

$$\text{Cov}(a_{R_1}, a_{R_2}) = -\frac{1}{N-1}\sigma_a^2.$$

Substitution of this back into (1) yields

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \left\{n - \frac{n(n-1)}{N-1}\right\}\sigma_a^2 = n\left(1 - \frac{n-1}{N-1}\right)\sigma_a^2.$$

Equivalently

$$\text{Var}(\bar{X}_n) = \frac{\sigma_a^2}{n}\left(1 - \frac{n-1}{N-1}\right).$$

In the case of sampling *with replacement* the variance of the sample mean is just σ_a^2/n . Thus the variance is smaller in the case of sampling without replacement by exactly the factor $1 - (n-1)/(N-1)$.

If $0 < \liminf(n/N) \leq \limsup(n/N) < 1$ and

$$\frac{\max_{1 \leq j \leq N} |a_j - \bar{a}_N|^2}{\sum_{j=1}^n (a_j - \bar{a}_N)^2} \rightarrow 0$$

then $(\bar{X}_n - \bar{a}_N)/\sqrt{\text{Var}(\bar{X}_n)} \rightarrow_d N(0, 1)$.

Do **either** problem 4 **or** problem 5.

4. (30 points). Suppose that X_1, X_2, \dots are i.i.d. $(0, 1)$ and $S_k \equiv X_1 + \dots + X_k$ for $k = 1, 2, 3, \dots$

(a) What is the limiting distribution of:

(i) $\max_{1 \leq k \leq n} n^{-1/2} S_k$?

(ii) $n^{-1} \sum_{k=1}^n 1_{[S_k > 0]}$?

(iii) $n^{-1} \sum_{k=[an]}^n 1_{[S_k > c\sqrt{k}]}$ where $a, c > 0$?

(iv) $n^{-1} \inf\{k \geq 1 : S_k \geq b\sqrt{n}\}$ for $b > 0$?

(b) Find a sequence of normalizing constants c_n so that

$$\frac{1}{c_n} \sum_{k=1}^n S_k^4 \rightarrow_d \text{something as } n \rightarrow \infty$$

and identify “something” in terms of BM

Solution: (a) Let $\mathbb{S}_n(t) \equiv n^{-1/2} \sum_{i=1}^{[nt]} X_i$ denote the partial sum process and let \mathbb{S} denote standard Brownian motion. Then:

(i) $\max_{1 \leq k \leq n} n^{-1/2} S_k = \max_{0 \leq t \leq 1} \mathbb{S}_n(t) \rightarrow_d \sup_{0 \leq t \leq 1} \mathbb{S}(t)$.

(ii) $n^{-1} \sum_{k=1}^n 1_{[S_k > 0]} = \int_0^1 1_{[\mathbb{S}_n(t) > 0]} dt \rightarrow_d \int_0^1 1_{[\mathbb{S}(t) > 0]} dt$ (since $g(x) \equiv \int_0^1 1_{[0, \infty)}(x(t)) dt$ for $x \in C[0, 1]$ is continuous a.s. $P^{\mathbb{S}}$).

(iii) In this case

$$n^{-1} \sum_{k=[an]}^n 1_{[S_k > c\sqrt{k}]} = \int_{[an]/n}^1 1_{[\mathbb{S}_n(t) > c\sqrt{t}]} dt \rightarrow_d \int_a^1 1_{[\mathbb{S}(t) > c\sqrt{t}]} dt.$$

(iv) Now we have

$$\begin{aligned} n^{-1} \inf\{k \geq 1 : S_k \geq b\sqrt{n}\} &= \inf\{k/n : \mathbb{S}_n(k/n) \geq b\} \\ &\rightarrow_d \inf\{t > 0 : \mathbb{S}(t) \geq b\} \end{aligned}$$

since $g(x) \equiv \inf\{t > 0 : g(t) \geq b\}$ for $x \in C[0, 1]$ is continuous a.s. $P^{\mathbb{S}}$.

(b) We can write, by choosing $c_n = n^3$,

$$\begin{aligned} \frac{1}{c_n} \sum_{k=1}^n S_k^4 &= \frac{1}{c_n} \sum_{k=1}^n n^2 \left(\frac{S_k}{\sqrt{n}} \right)^4 = \frac{n^2}{c_n} \sum_{k=1}^n \mathbb{S}_n^4(k/n) \\ &= \frac{1}{n} \sum_{k=1}^n \mathbb{S}_n^4(k/n) \text{ if } c_n = n^3 \\ &\rightarrow_d \int_0^1 \mathbb{S}^4(t) dt. \end{aligned}$$

5. (30 points). Suppose \mathbb{S} is standard Brownian motion, and $-a < 0 < b$. Let $\tau \equiv \tau_{a,b} \equiv \inf\{t > 0 : \mathbb{S}(t) \notin (-a, b)\}$.
- (a) Sketch the proof that

$$P(\mathbb{S}(\tau) = -a) = \frac{b}{b+a} \quad \text{and} \quad P(\mathbb{S}(\tau) = b) = \frac{a}{b+a}.$$

- (b) What is $E\mathbb{S}^2(\tau)$? What is $E\tau$?
What is the relationship between them?
- (c) How would you bound $E\tau^2$?

Solution: See PFS, Theorem 12.6.1, page 319, Exercise 12.7.3, page 325, and class notes.

Do **one** of problems 6, 7, **or** problem 8.

6. (36 points). Prove *one* of the following two inequalities:

$$P(|X| \geq 1/\epsilon) \leq \frac{7}{\epsilon} \int_0^\epsilon (1 - \text{Real}\phi(t)) dt;$$

$$P(|X| \geq 1/\epsilon) \leq \frac{1}{2\epsilon} \int_{|t| \leq 2\epsilon} |1 - \phi(t)| dt.$$

Hints: In proving the first inequality, you may use the fact that $\inf_{|y| \geq 1} (1 - \sin(y)/y) = (1 - \sin(1)) = .1585 \dots \geq 1/7$.

Solution: The first inequality is proved in Section 9.5, page 210, PFS. The second inequality is proved as follows: first note that for $T \in (0, \infty)$ we have, by Fubini's theorem,

$$\begin{aligned} \frac{1}{2T} \int_{-T}^T \phi(t) dt &= \frac{1}{2T} \int_{-T}^T E(\cos(tX) + i \sin(tX)) dt \\ &= \frac{1}{2T} E \left\{ \int_{-T}^T (\cos(tX) + i \sin(tX)) dt \right\} \\ &= E \left(\frac{\sin(TX)}{TX} \right). \end{aligned}$$

It follows that

$$\begin{aligned}
\left| \frac{1}{2T} \int_{-T}^T \phi(t) dt \right| &\leq E \left| \frac{\sin(TX)}{TX} \right| \\
&\leq E \left| \frac{\sin(TX)}{TX} \right| 1_{\{|X| \geq \epsilon\}} + E \left| \frac{\sin(TX)}{TX} \right| 1_{\{|X| < \epsilon\}} \\
&\leq \frac{1}{T\epsilon} P(|X| \geq \epsilon) + 1 - P(|X| \geq \epsilon)
\end{aligned}$$

since $|\sin(y)| \leq 1$ and $|\sin(y)/y| \leq 1$. Choosing $T = 2/\epsilon$ yields

$$\left| \frac{\epsilon}{4} \int_{-2/\epsilon}^{2/\epsilon} \phi(t) dt \right| \leq 1 - \frac{1}{2} P(|X| \geq \epsilon)$$

or, equivalently,

$$\begin{aligned}
P(|X| \geq \epsilon) &\leq 2 - \left| \frac{\epsilon}{2} \int_{-2/\epsilon}^{2/\epsilon} \phi(t) dt \right| \\
&= \frac{\epsilon}{2} \int_{|t| \leq 2/\epsilon} dt - \left| \frac{\epsilon}{2} \int_{-2/\epsilon}^{2/\epsilon} \phi(t) dt \right| \\
&\leq \frac{\epsilon}{2} \int_{-2/\epsilon}^{2/\epsilon} |1 - \phi(t)| dt;
\end{aligned}$$

i.e. (2) holds.

7. (36 points). Suppose that you are given the law of the iterated logarithm for Brownian motion \mathbb{S} at ∞ :

$$\limsup_{t \rightarrow \infty} \frac{\mathbb{S}(t)}{\sqrt{2t \log \log t}} = 1 \quad a.s. \quad (2)$$

(a) Prove the *time reversal* property of Brownian motion: if \mathbb{S} is standard Brownian motion, then the process $\tilde{\mathbb{S}}(t) \equiv t\mathbb{S}(1/t)$ is also standard Brownian motion.

(b) Use (a) together with (2) to prove the LIL for Brownian motion at 0:

$$\limsup_{t \rightarrow 0} \frac{\mathbb{S}(t)}{\sqrt{2t \log \log(1/t)}} = 1 \quad a.s. \quad (3)$$

Solution: (a) $\tilde{\mathbb{S}}(t) \equiv t\mathbb{S}(1/t)$ is clearly a mean zero Gaussian process because \mathbb{S} is. Furthermore for $0 \leq s, t < \infty$ we have

$$\begin{aligned} E(\tilde{\mathbb{S}}(s)\tilde{\mathbb{S}}(t)) &= E(st\mathbb{S}(1/s)\mathbb{S}(1/t)) \\ &= st \{(1/s) \wedge (1/t)\} \\ &= s \wedge t. \end{aligned}$$

Moreover $\tilde{\mathbb{S}}(0) = \lim_{t \searrow 0} t\mathbb{S}(1/t) = \lim_{t \nearrow \infty} \mathbb{S}(v)/v = 0$ a.s. by the strong law of large numbers. Thus $\tilde{\mathbb{S}}$ is standard Brownian motion.

(b) For $1/s = t$ we have

$$\begin{aligned} \frac{\mathbb{S}(t)}{\sqrt{2t \log \log t}} &= \frac{\mathbb{S}(1/s)}{\sqrt{2(1/s) \log \log(1/s)}} \\ &= \frac{s\mathbb{S}(1/s)}{\sqrt{2s \log \log(1/s)}} = \frac{\tilde{\mathbb{S}}(s)}{\sqrt{2s \log \log(1/s)}}. \end{aligned}$$

where $\mathbb{S}(t) \equiv t\mathbb{S}(1/t)$. It follows that

$$\begin{aligned} 1 &= \limsup_{t \rightarrow \infty} \frac{\mathbb{S}(t)}{\sqrt{2t \log \log t}} \\ &= \limsup_{s \rightarrow 0} \frac{\tilde{\mathbb{S}}(s)}{\sqrt{2s \log \log(1/s)}} = \limsup_{s \rightarrow 0} \frac{\mathbb{S}(s)}{\sqrt{2s \log \log(1/s)}} \end{aligned}$$

where the last line follows from the result of (a). Thus the LIL for Brownian motion at 0 follows from the LIL for Brownian motion at ∞ via time inversion.

8. (36 points). Suppose that \mathbb{S} is standard Brownian motion on $[0, \infty)$, and define $\mathbb{Y}(t) \equiv e^{-t}\mathbb{S}(e^{2t})$ for $t \in \mathbb{R}$.
- Compute $E(\mathbb{Y}(t))$ and $Var(\mathbb{Y}(t))$ for $t \in \mathbb{R}$.
 - Compute $Cov(\mathbb{Y}(s), \mathbb{Y}(t))$ for $s, t \in \mathbb{R}$.
 - What is the joint distribution of $(\mathbb{Y}(s), \mathbb{Y}(t))$?
 - Show that \mathbb{Y} is a stationary process.
 - Is there a connection between \mathbb{Y} and a Brownian bridge process \mathbb{U} (perhaps divided by $\sqrt{t(1-t)}$)?

Solution: (a) $E(\mathbb{Y}(t)) = e^{-t}E(\mathbb{S}(e^{2t})) = e^{-t} \cdot 0 = 0$ and

$$Var(\mathbb{Y}(t)) = e^{-2t}Var(\mathbb{S}(e^{2t})) = e^{-2t}e^{2t} = e^0 = 1$$

for all $t \in \mathbb{R}$.

(b) For $s, t \in \mathbb{R}$ we have

$$\begin{aligned} \text{Cov}(\mathbb{Y}(s), \mathbb{Y}(t)) &= E\{e^{-s}\mathbb{S}(e^{2s})e^{-t}\mathbb{S}(e^{2t})\} = e^{-(s+t)} \cdot e^{2s \wedge t} \\ &= \begin{cases} e^{-(s+t)}e^{2s}, & \text{if } s \leq t \\ e^{-(s+t)}e^{2t}, & \text{if } t \leq s, \end{cases} \\ &= e^{-|t-s|}. \end{aligned}$$

(c) The joint distribution of $(\mathbb{Y}(s), \mathbb{Y}(t))$ is bivariate normal with 0 means, variances 1, and correlation $e^{-|t-s|}$.

(d) Let t_1, \dots, t_k be points in \mathbb{R} , and let $h > 0$. Then the joint distribution of $(\mathbb{Y}(t_1 + h), \dots, \mathbb{Y}(t_k + h))$ is $N_k(\mathbf{0}, \Sigma)$ where $\Sigma = (\sigma_{i,j})_{i,j=1}^k$ is the variance-covariance matrix of $(\mathbb{Y}(t_1 + h), \dots, \mathbb{Y}(t_k + h))$. But this is the matrix with 1's on the diagonal and off-diagonal entries σ_{ij}

$$\exp(-|t_i + h - (t_j + h)|) = \exp(-|t_i - t_j|),$$

which is exactly the covariance matrix of $(\mathbb{Y}(t_1), \dots, \mathbb{Y}(t_k))$. Thus \mathbb{Y} is stationary.

(e) Yes: note that with $\mathbb{Z}(t) \equiv \mathbb{U}(t)/\sqrt{t(1-t)}$ we have

$$\begin{aligned} \mathbb{Z}(t) &= \frac{\mathbb{U}(t)}{\sqrt{t(1-t)}} \stackrel{d}{=} \frac{1}{\sqrt{t(1-t)}}(1-t)\mathbb{S}\left(\frac{t}{1-t}\right) \\ &= \sqrt{\frac{1-t}{t}}\mathbb{S}\left(\frac{t}{1-t}\right) = \mathbb{Y}\left(\frac{1}{2}\log\frac{t}{1-t}\right). \end{aligned}$$