Probability Theory
2015/16 Semester IIb
Instructor: Daniel Valesin
Final Exam
14/6/2016
Duration: 3 hours

Name: ______________________________
Student number: ____________

This exam contains 10 pages (including this cover page) and 8 problems. Enter all requested information on the top of this page, and put your initials on the top of every page, in case the pages become separated.

You are allowed to have two hand-written sheets of paper and a calculator.

You are required to show your work on each problem except for Problem 1 (True or False).

Do not write on the table below.

Problem	Points	Score
1	14	
2	14	
3	14	
4	14	
5	14	
6	7	
7	7	
8	6	
Total:	90	

Standard Normal cumulative distribution function
The value given in the table is $F_X(x)$ for $X \sim \mathcal{N}(0,1)$.

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

1. (a) (7 points) A closet has 5 pairs of shoes. I open the closet in the dark and take 4 shoes at random, without replacement (that is: whenever I take a shoe, it is chosen uniformly among all shoes that are still inside the closet). Find the probability that I can form at least one pair with the shoes I took.

   (b) (7 points) Bas is doing a cycling tour of Morocco. Today he wakes up in Rabat and needs to travel to Casablanca. Before he starts, he can study a map and ask for directions. If he does both, the probability that he will get lost is 0.15. If he studies the map but does not ask for directions, the probability that he will get lost is 0.4. If he asks for directions but does not study the map, the probability that he will get lost is 0.3. If he does neither, the probability that he will get lost is 0.7.
   Suppose that he uses a fair coin toss to decide whether or not to ask for directions (heads → asks; tails → does not ask), and another independent fair coin toss to decide whether or not to study the map (heads → studies; tails → does not study).
   Given that he arrives in Casablanca without getting lost, what is the probability that he studied the map?

Solution. (a) Without any restrictions, the number of ways to take 4 shoes out of the 10 available shoes is $\binom{10}{4}$. The number of ways to take 4 shoes out of the 10 available shoes so that no pairs can be formed is $\binom{5}{4} \cdot 2^4$ (we choose 4 distinct pairs and 1 shoe from each pair). Thus, the desired probability is

$$1 - \frac{\binom{5}{4} \cdot 2^4}{\binom{10}{4}} = \frac{13}{21}.$$

(b) Define the events

$$\begin{aligned} E_M &= \{\text{Bas studies his map}\} \\ E_D &= \{\text{Bas asks for directions}\} \\ L &= \{\text{Bas gets lost}\}. \end{aligned}$$

Then,

$$\begin{aligned} \mathbb{P}(E_M \mid L^c) &= \frac{\mathbb{P}(L^c \cap E_M \cap E_D) + \mathbb{P}(L^c \cap E_M \cap E_D^c)}{\mathbb{P}(L^c \cap E_M \cap E_D) + \mathbb{P}(L^c \cap E_M \cap E_D^c) + \mathbb{P}(L^c \cap E_M^c \cap E_D) + \mathbb{P}(L^c \cap E_M^c \cap E_D^c)} \\ &= \frac{\frac{1}{4} \cdot 0.85 + \frac{1}{4} \cdot 0.6}{\frac{1}{4} \cdot 0.85 + \frac{1}{4} \cdot 0.6 + \frac{1}{4} \cdot 0.7 + \frac{1}{4} \cdot 0.3} = 0.591837. \end{aligned}$$

2. Let $\ldots, X_{-3}, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots$ be independent Bernoulli(p) random variables, $p \in (0,1)$. Define

$$Y_n = -1 + \min\{m \geq n : X_m \neq X_n\} - \max\{m \leq n : X_m \neq X_n\}, \qquad n \in \{\ldots, -1, 0, 1, \ldots\}.$$

(a) (7 points) Find the probability mass function of Y_n.

(b) (7 points) Find $\mathbb{E}(Y_n)$. You may use the formula:

$$\sum_{n=1}^{\infty} n^2 \cdot q^n = \frac{q+q^2}{(1-q)^3}, \quad 0 < q < 1.$$

Solution. (a) For $k \in \{1, 2, \ldots\}$,

$$\begin{aligned}
f_{Y_n}(k) &= \mathbb{P}(Y_n = k) \\
&= \sum_{\ell=1}^{k} \mathbb{P}(X_{n-\ell} = 0,\ X_{n-\ell+1} = \cdots = X_{n-\ell+k} = 1,\ X_{n-\ell+k+1} = 0) \\
&\quad + \sum_{\ell=1}^{k} \mathbb{P}(X_{n-\ell} = 1,\ X_{n-\ell+1} = \cdots = X_{n-\ell+k} = 0,\ X_{n-\ell+k+1} = 1) \\
&= \sum_{\ell=1}^{k} (1-p)^2 p^k + \sum_{\ell=1}^{k} (1-p)^k p^2 \\
&= kp^k(1-p)^2 + k(1-p)^k p^2.
\end{aligned}$$

(b)

$$\begin{aligned}
\sum_{k=1}^{\infty} k \cdot f_{Y_n}(k) &= \sum_{k=1}^{\infty} k \cdot (kp^k(1-p)^2 + k(1-p)^k p^2) \\
&= (1-p)^2 \sum_{k=1}^{\infty} k^2 p^k + p^2 \sum_{k=1}^{\infty} k^2 (1-p)^k \\
&= (1-p)^2 \cdot \frac{p+p^2}{(1-p)^3} + p^2 \cdot \frac{1-p+(1-p)^2}{p^3} \\
&= \frac{5p^2 - 5p + 2}{p - p^2}.
\end{aligned}$$

3. We have n balls and m urns. The balls are placed in the urns one by one, and any given ball has the same probability of going into any urn. Let X be the number of urns that receive at least one ball.

   (a) (7 points) Find $\mathbb{E}(X)$.

   (b) (7 points) Find $\text{Var}(X)$.

Solution. For $i = 1, \ldots, m$, let X_i be the indicator function of the event that urn i receives at least one ball. Then,

$$\mathbb{E}(X_i) = \mathbb{P}(X_i = 1) = 1 - \mathbb{P}(X_i = 0) = 1 - \left(\frac{m-1}{m}\right)^n,$$

so

$$\text{Var}(X_i) = \left(1 - \left(\frac{m-1}{m}\right)^n\right) \cdot \left(\frac{m-1}{m}\right)^n.$$

Moreover, if $i \neq j$,

$$\text{Cov}(X_i, X_j) = \mathbb{E}(X_iX_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) = \mathbb{P}(X_i = X_j = 1) - \left(1 - \left(\frac{m-1}{m}\right)^n\right)^2.$$

Noting that

$$\begin{aligned}\mathbb{P}(X_i = X_j = 1) &= 1 - \mathbb{P}(\{X_i = 0\} \cup \{X_j = 0\}) \\ &= 1 - \mathbb{P}(\{X_i = 0\}) - \mathbb{P}(X_j = 0) + \mathbb{P}(X_i = X_j = 0) \\ &= 1 - 2\left(\frac{m-1}{m}\right)^n + \left(\frac{m-2}{m}\right)^n,\end{aligned}$$

we get the expression for the covariance:

$$\text{Cov}(X_i, X_j) = 1 - 2\left(\frac{m-1}{m}\right)^n + \left(\frac{m-2}{m}\right)^n - \left(1 - \left(\frac{m-1}{m}\right)^n\right)^2 = \left(\frac{m-2}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}.$$

Since $X = \sum_{i=1}^m X_i$, we thus obtain

$$\begin{aligned}\mathbb{E}(X) &= m\left(1 - \left(\frac{m-1}{m}\right)^n\right), \\ \text{Var}(X) &= m\text{Var}(X_1) + m(m-1)\text{Cov}(X_1, X_2) \\ &= m\left(1 - \left(\frac{m-1}{m}\right)^n\right) \cdot \left(\frac{m-1}{m}\right)^n \\ &\qquad + m(m-1)\left(\left(\frac{m-2}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}\right).\end{aligned}$$

4. (a) (7 points) A random variable X has pdf

$$f_X(x) = \frac{c}{x^2 - x}, \qquad 2 < x < 5.$$

Find c and $\mathbb{E}(\lfloor X \rfloor)$ ($\lfloor \cdot \rfloor$ denotes the *floor function*. For a non-negative real number x, $\lfloor x \rfloor$ is the largest integer n such that $n \leq x$).

(b) (7 points) Two random variables Y and Z have joint probability density function

$$f_{Y,Z}(y, z) = \frac{e^{-z}}{z}, \quad 0 < y < z < \infty.$$

Compute $\mathbb{E}(Y^2 \mid Z = z)$.

Solution. (a)

$$1 = c \cdot \int_2^5 \frac{1}{x^2 - x}\, dx = c \cdot \int_2^5 \left(\frac{1}{x-1} - \frac{1}{x} \right) = c \cdot \log\left(\tfrac{8}{5}\right) \Longrightarrow c = 1/\log\left(\tfrac{8}{5}\right)$$

$$\begin{aligned}
\mathbb{E}(\lfloor X \rfloor) &= \int_2^5 \lfloor x \rfloor \cdot f_X(x)\, dx \\
&= 2 \cdot \int_2^3 f_X(x)\, dx + 3 \cdot \int_3^4 f_X(x)\, dx + 4 \cdot \int_4^5 f_X(x)\, dx \\
&= 2 \cdot \log\left(\tfrac{4}{3}\right) + 3 \cdot \log\left(\tfrac{9}{8}\right) + 4 \cdot \log\left(\tfrac{16}{15}\right).
\end{aligned}$$

(b) For any $y > 0$,

$$f_{Y|Z}(y \mid z) = \frac{e^{-z}/z}{\int_0^z e^{-z}/z\, dy} = \frac{1}{z}, \quad y \in (0, z).$$

Hence,

$$\mathbb{E}(Y^2 \mid Z = z) = \int_0^z y^2 \cdot \frac{1}{z}\, dy = \frac{z^2}{3}.$$

5. (a) (7 points) Prove that, if $X \sim \text{Binomial}(n, p)$, then the moment-generating function of X is $M_X(t) = (1 + p(e^t - 1))^n$.

   (b) (7 points) Prove that, if $Y \sim \text{Poisson}(\lambda)$, then the moment-generating function of Y is $M_Y(t) = e^{\lambda(e^t - 1)}$.

Solution. (a)

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{i=0}^{n} e^{it} \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=0}^{n} \binom{n}{i} (pe^t)^i (1-p)^{n-i} = (1 - p + pe^t)^n.$$

(b)

$$M_Y(t) = \mathbb{E}(e^{tY}) = \sum_{i=0}^{\infty} e^{it} \frac{\lambda^i}{i!} e^{-\lambda} = e^{-\lambda} \cdot \frac{(\lambda e^t)^i}{i!} = e^{-\lambda} \cdot e^{\lambda e^t} = e^{\lambda(-1+e^t)}.$$

6. (7 points) If $X_1, X_2, \ldots$ are independent Poisson(1) random variables, then for each n,

$$X_1 + \cdots + X_n \sim \text{Poisson}(n)$$

(you do not need to prove this). Use this fact to prove that

$$\text{for all } \varepsilon > 0, \qquad \sum_{i \in \mathbb{N}:\, (1-\varepsilon)n \le i \le (1+\varepsilon)n} \frac{n^i}{i!} \cdot e^{-n} \xrightarrow{n\to\infty} 1$$

and

$$\sum_{i=0}^{n} \frac{n^i}{i!} \cdot e^{-n} \xrightarrow{n\to\infty} \frac{1}{2}.$$

Solution. Let $Y_n = \sum_{i=1}^n X_i \sim \text{Poisson}(n)$. Then, by the Weak Law of Large Numbers, we have

$$\frac{Y_n}{n} \xrightarrow[\mathbb{P}]{n\to\infty} \mathbb{E}(X_i) = 1,$$

which means that, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\left| \frac{Y_n}{n} - 1 \right| \le \varepsilon \right) \xrightarrow{n\to\infty} 1,$$

which means that

$$\mathbb{P}\left(Y_n \in ((1-\varepsilon)n, (1+\varepsilon)n)\right) \xrightarrow{n\to\infty} 1,$$

that is,

$$\sum_{i \in \mathbb{N}:\, (1-\varepsilon)n \le i \le (1+\varepsilon)n} f_{Y_n}(i) = \sum_{i \in \mathbb{N}:\, (1-\varepsilon)n \le i \le (1+\varepsilon)n} \frac{n^i}{i!} \cdot e^{-n} \xrightarrow{n\to\infty} 1.$$

Next, by the Central Limit Theorem, $\frac{Y_n - n}{\sqrt{n}} \xrightarrow[(\mathrm{d})]{n\to\infty} \mathcal{N}(0,1)$, so, letting $Z \sim \mathcal{N}(0,1)$,

$$\sum_{i=0}^{n} \frac{n^i}{i!} \cdot e^{-n} = \mathbb{P}\left(Y_n \le n\right) = \mathbb{P}\left(\frac{Y_n - n}{\sqrt{n}} \le 0 \right) \xrightarrow{n\to\infty} \mathbb{P}(Z \le 0) = \frac{1}{2}.$$

7. (7 points) Suppose $X_1, X_2, \ldots$ are Bernoulli(p) random variables and assume that, for some constant $K > 0$, we have

$$\mathrm{Cov}(X_i, X_j) = 0 \quad \text{for all } i, j \text{ with } |i - j| > K.$$

Show that $\frac{X_1 + \cdots + X_n}{n}$ converges in probability to p as $n \to \infty$.
Hint. First prove that $|\mathrm{Cov}(X_i, X_j)| \leq 1$ for every i, j. Then use this to obtain an upper bound for $\mathrm{Var}(\sum_{i=1}^n X_i)$. Finally, try to repeat the proof of the weak law or large numbers, using Chebychev's inequality.

Solution. We have $\mathrm{Var}(X_i) = p(1-p)$ for each i, so

$$|\mathrm{Cov}(X_i, X_j)| \leq \mathrm{Var}(X_i) \cdot \mathrm{Var}(X_j) = p^2(1-p)^2 \leq 1.$$

Then,

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{i=1}^n \sum_{j \in \{i+1,\ldots,n\}} \mathrm{Cov}(X_i, X_j) \leq np(1-p) + 2Kn.$$

For any $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{\sum_{i=1}^n X_i}{n} - p\right| > \varepsilon\right) = \mathbb{P}\left(\left(\sum_{i=1}^n X_i - pn\right)^2 > \varepsilon^2 n^2\right)$$

$$\leq \frac{\mathbb{E}\left((\sum_{i=1}^n X_i - pn)^2\right)}{\varepsilon^2 n^2} = \frac{\mathrm{Var}(\sum_{i=1}^n X_i)}{\varepsilon^2 n^2} \leq \frac{np(1-p) + 2Kn}{\varepsilon^2 n^2} \xrightarrow{n \to \infty} 0.$$

8. (6 points) A person has 100 light bulbs whose lifetimes (in hours) are independent random variables following an Exponential(5) distribution (that is, with pdf $f(x) = \frac{1}{5}e^{-x/5}$ for $x > 0$). If the bulbs are used one at a time, with a failed bulb being replaced immediately by a new one, approximate the probability that there is still a working bulb after 525 hours.

   Solution. Let X_i be the lifetime of the ith light bulb, then $X = \sum_{i=1}^{100} X_i$ has expectation $\mathbb{E}(X) = 500$ and $\mathrm{Var}(X) = 2500$, so that the Central Limit Theorem implies that

$$\mathbb{P}(X \geq 525) = \mathbb{P}\left(\frac{X - 500}{50} \geq \frac{525 - 500}{50}\right) \approx 1 - F_Z(1/2),$$

   where $Z \sim \mathcal{N}(0, 1)$. The approximate answer is thus $1 - 0.6915$.