Probability and Statistics
FS 2016
Final Exam
18.08.2016
Time Limit: 180 Minutes

Name: ____________________

Student ID: ____________________

This exam contains 18 pages (including this cover page) and 10 questions. Formulae sheet can be found at the end.

Please justify all your steps carefully. Otherwise no credit will be given.

Grade Table (for grading use only, please leave empty)

Question	Points	Score
1	7	
2	10	
3	7	
4	12	
5	10	
6	12	
7	10	
8	12	
9	10	
10	10	
Total:	100	

GOOD LUCK

1. (7 points) Let X and Y have a continuous distribution with joint p.d.f.

$$f(x,y) = \begin{cases} x+y & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 1\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Compute the covariance $\mathrm{Cov}(X,Y)$.

2. (10 points) Test for a certain disease has 99% accuracy, namely, the probability of a positive test is 0.99 when the tested person has the disease. Its false positive rating is 5%, i.e., the probability that a positive test is 0.05 if the tested person does not have the disease. Suppose one in ten thousand people in the age group of Mr. T. has this disease and suppose that Mr.T. is tested and the test result is positive. What is the *probability that Mr.T. has the disease*?

3. (7 points) Let X be a random variable with exponential distribution of parameter λ.

   (a) (4 points) *Show that X satisfies the memoryless property*, namely

   $$\mathbb{P}(X \geq t + h \,|\, X \geq h) = \mathbb{P}(X \geq t) \quad \forall\, t, h \geq 0.$$

   (b) (3 points) *Compute*

   $$\mathbb{E}\left[X \;|X \geq t\right].$$

4. (12 points) A random sample of n items is to be taken from a distribution with mean μ and standard deviation σ.

   (a) (5 points) Use the Chebyshev inequality to determine the *smallest number of items* n that must be taken in order to satisfy the following relation:

   $$\mathbb{P}\left(|\overline{X}_n - \mu| \leq \frac{\sigma}{4}\right) \geq 0.99 \quad \text{where} \quad \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

   (b) (7 points) Use the *central limit theorem* to determine the *smallest number* of items n that must be taken in order to satisfy the relation in part (a) approximately.

5. (10 points) (The weak law of large numbers.)

   Let $(X_j)_{j\geq 1}$ be a sequence of random variables such that

   $$\mathbb{E}(X_j) = 0, \quad \text{and} \quad \mathbb{E}(X_j^2) \leq c^* < \infty \quad \forall j = 1, 2 \ldots .$$

   Assume further that $\mathbb{E}(X_j X_k) = 0$ if $j \neq k$.

   Prove that $\overline{X}_n := (\sum_{i=1}^n X_i)/n$ *converges to zero in probability as* n *tends to infinity.*

6. (12 points) Suppose that $X_1, ..., X_n$ form an i.i.d random sample from Gaussian distribution with unknown mean μ and variance σ^2. We know, however, that mean and the standard deviation have the same value, i.e.,

$$\mu = \sigma =: \theta.$$

   (a) (10 points) *Compute the Maximum Likelihood Estimator* $\theta_{MLE}^{(n)}$ of the common value of the mean and the standard deviation.

   (b) (2 points) *Show that as n tends to infinity, $\theta_{MLE}^{(n)}$ converges almost surely to the true value of θ.*

7. (10 points) Suppose that $\mathbf{X} = (X_1, ..., X_n)$ form an i.i.d. random sample from the Bernoulli distribution with parameter θ, which is unknown ($0 < \theta < 1$). Suppose also that the prior distribution of θ is the Beta distribution with parameters $\alpha > 0$ and $\beta > 0$. *Prove that the posterior distribution of θ given that $\mathbf{x} = (x_1, \ldots, x_n)$ is the Beta distribution with parameters $\alpha + \sum_{i=1}^n x_i$ and $\beta + n - \sum_{i=1}^n x_i$.*

8. (12 points) Suppose that a random sample $\mathbf{X} = (X_1, \ldots, X_n)$ of size n is taken from a Poisson distribution for which the value of the mean θ is unknown, and the prior distribution of θ is a Gamma distribution with parameters $\alpha, \beta > 0$ for which the mean is μ_0, i.e. by the expected value formula from item (32) of the formula sheet, $\mu_0 = \alpha/\beta$.

   (a) (8 points) Show that the *posterior distribution of θ is again Gamma with parameters $\alpha' = \alpha + \sum_{i=1}^n x_i$ and $\beta' = \beta + n$.*

   (b) (4 points) *Show that there is a constant γ_n such that*

   $$\mathbb{E}[\,\theta \mid \mathbf{X}\,] = (\text{the mean of the posterior distribution of } \theta) = \gamma_n \overline{X}_n + (1 - \gamma_n)\mu_0,$$

   where $\overline{X}_n = (\sum_{i=1}^n X_i)/n$. Also *show that*

   $$\lim_{n \to \infty} \gamma_n = 1.$$

   Hint: You may use the expected value formula from item (32).

9. (10 points) Suppose that $X_1, ..., X_{100}$ form an i.i.d random sample from the normal distribution with *unknown mean* μ *and known variance 1*, and it is desired to test the following hypotheses $\mu_0 = 0.5$, i.e.,

$$H_0 : \quad \mu = \mu_0 \,, \qquad H_1 : \quad \mu \neq \mu_0 \,.$$

Consider a rejection region,

$$\mathcal{R} = \left\{ \overline{X}_n \notin [c_1, c_2] \right\} ,$$

where $\overline{X}_n = (\sum_{i=1}^{n} X_i)/n$ and the constants $c_1 < c_2$ to be determined by you. We reject the hypothesis H_0 if the observation is in the rejection region $\mathcal{R}$. Let

$$\pi(\mu) := \mathbb{P}(\mathcal{R} \mid \mu) = \text{probability of rejection if the mean is equal to } \mu.$$

(a) (5 points) *Derive a formula* for $\pi(\mu)$ in terms of c_1, c_2, μ and the error function Φ given on the first page of the formula sheet.

(b) (5 points) *Determine the values of the constants* $c_1 < c_2$ so that $\pi(0.5) = 0.10$ and the function $\pi(\mu)$ is *symmetric with respect around the point* $\mu = 0.5$.

10. (10 points) Let $(A_n)_{n\geq 1}$ be a sequence of independent events in a probability space. Suppose

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty\,.$$

Prove that

$$\mathbb{P}\big(\bigcap_{k\in\mathbb{N}} \bigcup_{n\geq k} A_n \big) = 1\,.$$

(This is known as the second Borel-Cantelli lemma. In the proof you may use the fact that $1 + x \leq e^x$ for all $x \in \mathbb{R}$.)

FORMULAE SHEET

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-z^2/2}\, dz, \quad x \in \mathbb{R}.$$

Let Φ^{-1} be its inverse function. Then,

$$\Phi^{-1}(0.1) = -1.28, \quad \Phi^{-1}(0.05) = -1.6449, \quad \Phi^{-1}(0.01) = -2.329.$$

1. A probability space $(\Omega, \mathcal{F})$ is a set Ω and a sigma-algebra $\mathcal{F}$ of subsets of Ω. An element A of $\mathcal{F}$ is called an *event*. A probability measure

$$\mathbb{P} : \mathcal{F} \to [0, 1],$$

is a a countably additive, measure.

2. For any disjoint events $\{A_i\}_i$ (i.e., $A_i \cap A_j = \emptyset$ whenever $i \neq j$),

$$\mathbb{P}\left(\cup_{i=1}^{d} A_i\right) = \sum_{i=1}^{d} \mathbb{P}(A_i).$$

3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.

4. Permutations. Selection of distinct objects with order but without replacement.

$$P_{n,k} = \frac{n!}{(n-k)!}, \quad 0 \leq k \leq n.$$

5. Combinations. Selection of distinct objects without order and replacement.

$$C_{n,k} = \frac{n!}{k!\,(n-k)!}, \quad 0 \leq k \leq n.$$

6. Conditional Probability. A and B are two events with $\mathbb{P}(B) > 0$.

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

7. Two events A and B are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$.

8. Total Probability Formula. For any partition $\{B_i\}_i$ (i.e., $B_i \cap B_j = \emptyset$ whenever $i \neq j$ and $\cup_i B_i = \Omega$),

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \mid B_i)\ \mathbb{P}(B_i).$$

9. Bayes Formula. For a partition $\{B_i\}_{i=1}^d$, (see item 8), an event A and an integer $i \in \{1, \ldots, d\}$,
$$\mathbb{P}\left(B_i \mid A\right) = \frac{\mathbb{P}\left(A \mid B_i\right)\ \mathbb{P}(B_i)}{\mathbb{P}\left(A\right)},$$
and $\mathbb{P}(A)$ is calculated by item 8.

10. Random Variable. r.v. $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ is measurable, where $\mathcal{B}$ is the set of all Borel subsets of $\mathbb{R}$.

11. Bernoulli Distribution with success parameter $p \in (0, 1)$. $X \in \{0, 1\}$ and
$$\mathbb{P}(X = 1) = p, \quad \mathbb{E}(X) = p, \quad Var(X) = p(1 - p).$$

12. Binomial Distribution with n trials and success parameter $p \in (0, 1)$. $X \in \{0, 1, \ldots, n\}$
$$\mathbb{P}(X = k) = C_{n,k}\ p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots n,$$
$$\mathbb{E}(X) = np, \quad Var(X) = np(1 - p).$$

13. Poisson Distribution with parameter $\lambda > 0$. $X \in \{0, 1, \ldots\}$
$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!}\ e^{-\lambda}, \quad k = 0, 1, \ldots,$$
$$\mathbb{E}(X) = \lambda, \quad Var(X) = \lambda.$$

14. Infinitely often, i.o. For a given countable sequence of events $\{A_i\}_i$, the set $\{A_i\ i.o.\}$ is defined by
$$\{A_i\ i.o.\} := \cap_{n=1}^{\infty}\ \cup_{m=n}^{\infty}\ A_m.$$

15. Borel-Cantelli Lemma 1. Suppose that $\{A_i\}_i$ satisfy
$$\sum_i \mathbb{P}(A_i) < \infty.$$
Then, $\mathbb{P}(\{A_i\ i.o.\}) = 0$.

16. Borel-Cantelli Lemma 2. Suppose that $\{A_i\}_i$ are mutually independent and satisfy
$$\sum_i \mathbb{P}(A_i) = \infty.$$
Then, $\mathbb{P}(\{A_i\ i.o.\}) = 1$.

17. Cumulative Distribution Function c.d.f. For a r.v. X,
$$F_X(x) := \mathbb{P}(X \leq x) = \mathbb{P}\left(\{\omega \in \Omega\ :\ X(\omega) \leq x\}\right), \quad x \in \mathbb{R}.$$
A c.d.f. $F_X(x) \in [0, 1]$ and it is non-decreasing, right continuous with left limits.

18. Probability Density Function p.d.f.

$$f_X(x) := \frac{d}{dx} F_X(x), \quad \text{whenever it exists.}$$

A p.d.f. f_X is non-negative and

$$\mathbb{P}(X \in A) = \int_A f_X(x)\; dx, \quad A \in \mathcal{B}.$$

In particular, $\int_{\mathbb{R}} f_X(x)\; dx = 1$.

19. Joint c.d.f. For two random variables X, Y,

$$F_{X,Y}(x,y) := \mathbb{P}(X \le x \text{ and } Y \le y), \quad x, y \in \mathbb{R}.$$

Then,

$$F_X(x) = \lim_{y \uparrow \infty} F_{X,Y}(x,y), \quad \text{and} \quad F_Y(y) = \lim_{x \uparrow \infty} F_{X,Y}(x,y).$$

20. Joint p.d.f. For two random variables X, Y,

$$f_{X,Y}(x,y) := \frac{\partial^2}{\partial x\; \partial y} F_{X,Y}(x,y) \quad \text{whenever exists.}$$

Then,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\; dy, \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\; dx.$$

21. Conditional p.d.f. For two random variables X, Y,

$$f_{X|Y}(x|y) := \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

22. Independence. Two random variables X, Y are *independent* (denoted by $X \perp Y$) if for every Borel sets A, B, the events $\{X \in A\}$ and $\{Y \in B\}$ are independent.

   $X \perp Y$ if and only if $F_{X,Y}(x,y) = F_X(x)\; F_Y(y)$ for every $x, y \in \mathbb{R}$.

   When a joint p.d.f. exists, $X \perp Y$ if and only if $f_{X,Y}(x,y) = f_X(x)\; f_Y(y)$ for every $x, y \in \mathbb{R}$.

23. Continuous Bayes Theorem. For two random variables X, Θ with a joint p.d.f.,

$$f_{\Theta|X}(\theta \mid x) = \frac{f_{X|\Theta}(x \mid \theta)\; f_\Theta(\theta)}{f_X(x)},$$

and $f_X(x)$ is computed by,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,\Theta}(x,\theta)\; d\theta = \int_{-\infty}^{\infty} f_{X|\Theta}(x|\theta)\; f_\Theta(\theta)\; d\theta.$$

24. Expected Value. X is an r.v.,

$$\mathbb{E}(X) := \int_\Omega X(\omega)\ d\mathbb{P}(\omega), \quad \text{(general)},$$
$$\mathbb{E}(X) := \sum_i\ x_i\ \mathbb{P}(X = x_i), \quad \text{(discrete)},$$
$$\mathbb{E}(X) := \int_\mathbb{R} x\ f_X(x)\ dx, \quad \text{(when p.d.f. exists)}.$$

For a "nice" function g (all bounded, measurable and more),

$$\mathbb{E}[g(X)] = \int_\Omega g(X(\omega))\ d\mathbb{P}(\omega), \quad \text{(general)},$$
$$\mathbb{E}[g(X)] = \sum_i\ g(x_i)\ \mathbb{P}(X = x_i), \quad \text{(discrete)},$$
$$\mathbb{E}[g(X)] = \int_\mathbb{R} g(x)\ f_X(x)\ dx, \quad \text{(when p.d.f. exists)}.$$

25. Conditional Expected Value. X, Y r.v.'s with a joint p.d.f., for a nice g,

$$\mathbb{E}\left[g(X) \mid Y = y\right] := \int_{-\infty}^{\infty} g(x)\ f_{X|Y}(x|y) dx.$$

26. Variance and Standard Deviation

$$\text{variance of X} = Var(X) := \mathbb{E}[X^2] - [\mathbb{E}(X)]^2 = \mathbb{E}\left[(X - E[X])^2\right],$$
$$\text{standard deviation of X} = \sigma_X := \sqrt{Var(X)}.$$

27. Covariance and Correlation

$$\text{covariance of X and Y} = Cov(X, Y) := \mathbb{E}\left[(X - E[X])\ (Y - E[Y])\right],$$
$$\text{correlation of X and Y} = Corr(X, Y) := \frac{Cov(X, Y)}{\sigma_X\ \sigma_Y} \in [-1, 1].$$

We have

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y).$$

28. Jensen's Inequality. For a convex function $\Phi : \mathbb{R} \to \mathbb{R}$,

$$\Phi\left(\mathbb{E}[X]\right) \le \mathbb{E}[\Phi(X)].$$

29. Moment Generating Function. For a r.v., X,

$$\Psi(t) := \mathbb{E}\left[e^{tX}\right], \quad t \in \mathbb{R}.$$

$\Psi(t)$ could be $+\infty$. When it is finite near the origin,

$$\mathbb{E}\left[\ X^k\ \right] = \frac{d^k}{dt^k}\Psi(0).$$

30. Exponential Distribution with parameter $\lambda > 0$. $X \in \mathbb{R}_+ := [0, \infty)$,

$$f_X(x) = \lambda\, e^{-\lambda x}, \quad x \geq 0,$$

and $f_X(x) = 0$ for all $x < 0$ and

$$\mathbb{E}(X) = \frac{1}{\lambda}, \quad Var(X) = \frac{1}{\lambda^2}.$$

31. Gaussian Distribution with mean μ and variance σ^2. $X \in \mathbb{R}$,

$$f_X(x) := \frac{1}{\sqrt{2\pi}\ \sigma}\ \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right), \quad x \in \mathbb{R}.$$

Denoted by $X \sim \mathcal{N}(\mu, \sigma^2)$. Then,

$$X \sim \mathcal{N}(\mu, \sigma^2) \quad \Leftrightarrow \quad Z := \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1).$$

$\mathcal{N}(0,1)$ is called the standard normal (or Gaussian).

32. Gamma distribution with parameters α, β. $X \in \mathbb{R}_+ := [0, \infty)$,

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}\ x^{\alpha-1}\ e^{-\beta x}, \quad x \geq 0,$$

and $f_X(x) = 0$ for all $x \leq 0$. Here $\Gamma(\alpha)$ is the Gamma function and for integer values $\Gamma(m) = (m-1)!$.

$$\mathbb{E}(X) = \frac{\alpha}{\beta}, \quad Var(X) = \frac{\alpha}{\beta^2}.$$

33. Beta distribution with parameters α, β. $\Theta \in [0,1]$,

$$f_\Theta(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\ \theta^{\alpha-1}\ (1-\theta)^{\beta-1}, \quad \theta \in [0,1],$$

and $f_\Theta(\theta) = 0$ for all $\theta \notin [0,1]$.

$$\mathbb{E}(\Theta) = \frac{\alpha}{\alpha+\beta}, \quad Var(\Theta) = \frac{\alpha\beta}{(\alpha+\beta)^2\ (1+\alpha+\beta)}.$$

34. Chebyshev's Inequality. For a r.v., X and a increasing function G, a real number a with $G(a) > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}\left[G(X)\right]}{G(a)}\ .$$

35. Law of Large Numbers. Let $\{X_i\}_i$ be an i.i.d. sequence. Set $\mu := \mathbb{E}[X_i]$, $\sigma^2 := Var(X_i)$ for any i,

$$\overline{X_n} := \frac{1}{n} \sum_{i=1}^{n} X_i, \quad Z_n := \frac{\sqrt{n}}{\sigma} \, [\overline{X_n} - \mu].$$

Then, weak law of large numbers state that $\overline{X_n}$ converges to μ in probability and strong law states that the convergence is almost surely.

36. Central Limit Theorem

The distribution of Z_n converges to the standard Gaussian, i.e., for every continuous and bounded function φ,

$$\lim_{n\to\infty} \mathbb{E}\left[\varphi(Z_n)\right] = \frac{1}{\sqrt{2\pi}} \int \varphi(x) \exp\left(-\frac{x^2}{2}\right) dx.$$

37. χ-Square (χ^2) Distribution.

χ^2 distribution with m degrees of freedom is the Gamma distribution with parameters $(m/2, 1/2)$. Denoted by $\chi^2(m)$. In particular,

$$X \sim \mathcal{N}(0,1) \quad \Rightarrow \quad X^2 \sim \chi^2(1),$$

$$X_i \sim \mathcal{N}(0,1) \text{ i.i.d} \quad \Rightarrow \quad \sum_{i=1}^{m} X_i^2 \sim \chi^2(m),$$

38. Student t distribution.

If $Z \sim \mathcal{N}(0,1)$, $Y \sim \chi^2(m)$, $Z \perp Y$, then,

$$T := \frac{Z}{\sqrt{Y/m}},$$

has t distribution with m degrees of freedom. Denoted by $t(m)$.

Its density is given by

$$f_T(t) = \frac{\Gamma((m+1)/2)}{\sqrt{m\pi}\,\Gamma(m/2)} \left(1 + \frac{t^2}{m}\right)^{-(m+1)/2}, \quad t \in\in \mathbb{R}.$$

39. If $\{X_i\}_{i=1}^n$ are i.i.d. with $\mathcal{N}(\mu, \sigma^2)$ distribution. Let $\overline{X_n}$ as in item 35. Set

$$S_n^2 := \sum_{i=1}^{n} (X_i - \overline{X_n})^2, \quad \sigma' := \left[\frac{1}{n-1} S_n^2\right]^{1/2}.$$

Then, $\overline{X_n} \perp S_n^2$ and

$$U := \frac{\sqrt{n}}{\sigma'} \, [\overline{X_n} - \mu]$$

has $t(n-1)$ distribution.

40. Two sided Confidence Interval. With the notation of item 39,

$$A := \overline{X_n} - T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right)\ \frac{\sigma'}{\sqrt{n}},$$
$$B := \overline{X_n} + T_{n-1}^{-1}\left(\frac{1+\gamma}{2}\right)\ \frac{\sigma'}{\sqrt{n}},$$

where T_m is the c.d.f. of the t distribution with m degrees of freedom, T_m^{-1} is its inverse function, parameter $\gamma \in (0,1)$ is the confidence level.

The interval (A, B) is γ-level two sided confidence interval.

41. One sided Confidence Interval With the notation of item 39,

$$\hat{A} := \overline{X_n} - T_{n-1}^{-1}(\gamma)\ \frac{\sigma'}{\sqrt{n}},$$
$$\hat{B} := \overline{X_n} + T_{n-1}^{-1}(\gamma)\ \frac{\sigma'}{\sqrt{n}}.$$

The interval $(\hat{A}, \infty)$ is a γ-level upper confidence interval.

The interval $(-\infty, \hat{B})$ is a γ-level lower confidence interval.

42. Change of Variables Formula Let $X = (X_1, \ldots, X_n)$ have a continuous joint distribution $f_X(x)$ for $x \in \mathbb{R}^n$. Let $Y = AX$ for some non-singular square matrix A. Then, the probability distribution function of Y is given by,

$$f_Y(y) = \frac{1}{|det(A)|}\ f_X(A^{-1}y), \quad y \in \mathbb{R}^n.$$