Probability and Statistics
FS 2016
Second Session Exam
07.02.2017
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	10	
2	12	
3	10	
4	12	
5	10	
6	8	
7	10	
8	12	
9	10	
10	6	
Total:	100	

GOOD LUCK

1. (10 points) Let X and Y have a continuous distribution with joint p.d.f.

$$f(x,y) = c\begin{cases} 2x+y & \text{for } 0 \le x \le 1 \text{ and } 0 \le y \le 2\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

for some constant c. *Compute the value of* c *and the covariance* $\mathrm{Cov}(X,Y)$.

Solution

The marginal density function of X is

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,\mathrm{d}y = c\begin{cases} 4x+2 & \text{for } 0 \le x \le 1\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Since the constant c making f_X a p.d.f. should satisfies

$$\int_{\mathbb{R}} f_X(x)\mathrm{d}x = 1.$$

We have $c = 1/4$.

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x)\,\mathrm{d}x = c\int_0^1 4x^2 + 2x\,\mathrm{d}x = \frac{7}{12}\,.$$

Similarly we have

$$f_Y(y) = \int_{\mathbb{R}} f(x,y)\,\mathrm{d}x = c\begin{cases} y+1 & \text{for } 0 \le y \le 2\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Hence

$$\mathbb{E}[Y] = \int_{\mathbb{R}} y f_Y(y)\,\mathrm{d}x = c\int_0^2 y^2 + y\,\mathrm{d}y = \frac{7}{6}\,.$$

Also,

$$\begin{aligned} \mathbb{E}[XY] &= \int_{\mathbb{R}} xyf(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= c\int_0^2\int_0^1 2x^2y + xy^2\,\mathrm{d}x\,\mathrm{d}y \\ &= c\int_0^2 \frac{2y}{3} + \frac{y^2}{2}\mathrm{d}y \\ &= \frac{2}{3}. \end{aligned}$$

Hence

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = -\frac{1}{72}\,.$$

2. (12 points) We are interested in studying the probability of success of a student at an entrance exam to two departments of a university. Consider the following events

$$
\begin{aligned}
A &= \{\text{The student is man}\},\\
A^c &= \{\text{The student is woman}\},\\
B &= \{\text{The student applied for department I}\},\\
B^c &= \{\text{The student applied for department II}\},\\
C &= \{\text{The student was accepted}\},\\
C^c &= \{\text{The student wasn't accepted}\}.
\end{aligned}
$$

We assume that we have the following probabilities (Berkeley 1973):

$\mathbb{P}(A) = 0.73$,
$\mathbb{P}(B \mid A) = 0.69$, $\mathbb{P}(B \mid A^c) = 0.24$,
$\mathbb{P}(C \mid A \cap B) = 0.62$, $\mathbb{P}(C \mid A^c \cap B) = 0.82$, $\mathbb{P}(C \mid A \cap B^c) = 0.06$, $\mathbb{P}(C \mid A^c \cap B^c) = 0.07$.

(a) (9 points) *Compute the probabilities* $\mathbb{P}[C|A]$, $\mathbb{P}[C|A^c]$.

(b) (3 points) *Do you think that in this examination women are disadvantaged?*

Solution

(a) Just computing

$$
\begin{aligned}
\mathbb{P}[C|A^c] &= \frac{\mathbb{P}[C \cap A^c]}{\mathbb{P}[A^c]} = \frac{\mathbb{P}[C \cap A^c \cap B] + \mathbb{P}[C \cap A^c \cap B^c]}{\mathbb{P}[A^c]}\\
&= \frac{0.82 \cdot 0.27 \cdot 0.24 + 0.07 \cdot 0.27 \cdot 0.76}{0.27} \sim 0.25,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{P}[C|A] &= \frac{\mathbb{P}[C \cap A]}{\mathbb{P}[A^c]} = \frac{\mathbb{P}[C \cap A \cap B] + \mathbb{P}[C \cap A \cap B^c]}{\mathbb{P}[A]}\\
&= \frac{0.62 \cdot 0.69 \cdot 0.73 + 0.06 \cdot 0.31 \cdot 0.73}{0.73} \sim 0.45.
\end{aligned}
$$

This shows that the percentage of women accepted are less than that of the men.

(b) The probability of being accepted, given than you are a woman who postulated at the department I is $P(C \mid A^c \cap B) = 0.82$. That value is bigger than the probability of being accepted, given than you are a man who postulated at the department I, $\mathbb{P}(C \mid A \cap B) = 0.62$. This indicates that in department I females are not disadvantaged.

The probability of being accepted, given than you are a woman who postulated to the department is $P(C \mid A^c \cap B^c) = 0.07$. This value is bigger than the probability of being accepted given than you are a man who postulated at the department II $\mathbb{P}(C \mid A \cap B^c) = 0.06$. This indicates that in department II females are neither disadvantaged.

The result of (a) is not explained by the gender, but much more but the fact that women apply to the department with bigger rejection rate.

3. (10 points) Let X be a random variable with uniform distribution on $[0,a]$.

   (a) (6 points) *Compute*
   $$\mathbb{P}(X \geq c \,|\, X \geq b) \quad \text{for } b < c < a.$$

   (b) (4 points) *Compute*
   $$\mathbb{E}\left[X \;|X \geq b\right].$$

Solution

(a) By definition of conditional probability,
$$\mathbb{P}(X \geq c \,|\, X \geq b) = \frac{\mathbb{P}(X \geq c, X \geq b)}{\mathbb{P}(X \geq b)} = \frac{\mathbb{P}(X \geq c)}{\mathbb{P}(X \geq b)} = \frac{a-c}{a-b}.$$

Thus conditioning on $X \geq b$, X is distributed as the uniform random variable on $[b,a]$.

(b) The uniform random variable on $[b,a]$ has expectation $(a+b)/2$, thus
$$\mathbb{E}\left[X \;|X \geq b\right] = (a+b)/2.$$

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:

$$\mathbb{P}\Big(|\overline{X}_n - \mu| \le \frac{\sigma}{3}\Big) \ge 0.90 \quad \text{where} \quad \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

   (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.

Solution

(a) By taking $G(x) = x^2$ in Chebyshev inequality we have

$$\mathbb{P}\Big(|\overline{X}_n - \mu| > \frac{\sigma}{3}\Big) \le \frac{\mathbb{E}[(\overline{X}_n - \mu)^2]}{(\sigma/3)^2} = \frac{\mathbb{E}[(\sum_{i=1}^n (X_i - \mu))^2]}{n^2(\sigma/3)^2} = \frac{\sum_{i=1}^n \mathbb{E}[(X_i - \mu)^2]}{n^2(\sigma/3)^2} = \frac{9}{n},$$

where in the third step we used the fact that $X_1, ..., X_n$ are independent. This implies

$$\mathbb{P}\Big(|\overline{X}_n - \mu| \le \frac{\sigma}{3}\Big) \ge 1 - \frac{9}{n}.$$

Thus in order to satisfy $\mathbb{P}(|\overline{X}_n - \mu| \le \sigma/3) \ge 0.90$, it suffice to have $1 - 9/n \ge 0.90$, which is $n \ge 90$.

(b) Let us denote $Z_n = \sqrt{n}(\overline{X}_n - \mu)/\sigma$. By central limit theorem, the random variable Z_n can be approximated by the standard normal distribution. Let Φ be the c.d.f. of the standard normal distribution. We have

$$\mathbb{P}\Big(|\overline{X}_n - \mu| \le \frac{\sigma}{3}\Big) = \mathbb{P}\Big(|Z_n| \le \frac{\sqrt{n}}{3}\Big) \approx \Phi\Big(\frac{\sqrt{n}}{3}\Big) - \Phi\Big(-\frac{\sqrt{n}}{3}\Big) = 2\Phi\Big(\frac{\sqrt{n}}{3}\Big) - 1.$$

Thus we need $\Phi\big(\sqrt{n}/3\big) \ge 0.95$, which implies $n \ge 25$.

5. (10 points) Let $(X_j)_{j\geq 1}$ be a sequence of independent (not identical) random variables

$$\mu_j := \mathbb{E}(X_j).$$

Assume that there exists $c^* > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \mu_j = \mu \quad \text{and} \quad \mathbb{E}(X_j - \mu_j)^2 = \mathrm{Var}(X_j) \leq c^*, \quad \forall j.$$

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

Solution

Note that $\mathbb{E}(\overline{X}_n) = \frac{1}{n}\sum_{j=1}^n \mu_j$. By independence of (X_i),

$$\mathrm{Var}(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) \leq \frac{c^*}{n}.$$

First let us note that $\mathbb{P}(|\mathbb{E}\left(\overline{X}_n\right) - \mu| \geq \epsilon/2) = 0$ for $n \geq N$, then for all $n \geq N$ Chebyshev's and triangular inequality give that for every $\varepsilon > 0$.

$$\mathbb{P}(|\overline{X}_n - \mu| \geq \varepsilon) \leq \mathbb{P}(|\overline{X}_n - \mathbb{E}\left(\overline{X}_n\right)| \geq \epsilon/2) + \mathbb{P}(|\mathbb{E}\left(\overline{X}_n\right) - \mu| \geq \epsilon/2) \leq \frac{\mathrm{Var}(\overline{X}_n)}{\varepsilon^2} \to 0.$$

6. (8 points) A gas station estimates that it takes at least λ minutes for a change of oil. The actual time varies from costumer to costumer. However, one can assume that this time will be well represented by an exponential random variable. The random variable X, therefore, possess the following density function

$$f(t) = e^{\lambda - t}\mathbf{1}_{\{t \geq \lambda\}},$$

i.e. $X = \lambda + Z$ where $Z \sim Exp(1)$. The following values were recorded from 10 clients randomly selected (the time is in minutes):

$$4.2,\ 3.1,\ 3.6,\ 4.5,\ 5.1.$$

Estimate the parameter λ using the estimator of maximum likelihood.

Solution

We have that the likelihood function is given by:

$$\begin{aligned} L(X_1, ..., X_n, \lambda) &= \prod_{i=1}^{n} \exp(\lambda - X_i)\mathbf{1}_{\{X_i \geq \lambda\}}, \\ &= \exp(n\lambda - \sum_{i=1}^{n} X_i)\mathbf{1}_{\{X_i \geq \lambda, \forall i\}}, \end{aligned}$$

we note that $f(\lambda) := \exp(n\lambda - \sum_{i=1}^{n} X_i) > 0$ is increasing, so its maximum is attained at the maximum point where $\mathbf{1}_{\{X_i \geq \lambda, \forall i\}} = 1$. Then the point that maximizes the likelihood is in $\lambda = \min_{i=1,..,n}\{X_i\} = 3.1$.

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* $\beta + n - \sum_{i=1}^n x_i$.

   Solution

   First we calculate the joint p.f. of $X_1, \cdots, X_n, \Theta$:

$$\begin{aligned} f_{X_1,\cdots,X_n,\Theta}(x_1, \cdots, x_n, \theta) &\propto \ \theta^y (1-\theta)^{n-y} \theta^{\alpha-1} (1-\theta)^{\beta-1} \\ &= \theta^{\alpha+y-1} (1-\theta)^{\beta+n-y-1} \end{aligned}$$

   where $y = x_1 + \cdots + x_n$. So conditioning on $X_1, \cdots, X_n$ at $(x_1, \cdots, x_n)$, Θ has the p.d.f.

$$f_{\theta|x_1,\cdots,x_n}(\theta) \propto \frac{\theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}}{f_{X_1,\cdots,X_n}(x_1, \cdots, x_n)} \propto \theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1},$$

   where $f_{X_1,\cdots,X_n}$ is the marginal p.d.f of $X_1, \cdots, X_n$ which does not depend on θ (by definition). We see that the posterior distribution of Θ is the Beta distribution with parameters $\alpha + y$ and $\beta + n - y$.

8. (12 points) We have 2 urns with red and white balls inside. The urn $i \in \{1,2\}$ has i red balls and $3-i$ white ones. We uniformly select an urn and extract with replacement n times. Define:

$$X_j := \begin{cases} 1 & \text{If the } j\text{-th ball is red,} \\ 0 & \text{If the } j\text{-th ball is white.} \end{cases}$$

We are interested in the following problem " Given that you see $(X_j)_{j=1}^n$, can you say from which urn the balls where taken?"

(a) (6 points) *Compute* $\mathbb{P}(X_1 = 1, X_2 = 1)$. *Are* X_1, X_2 *independent?*

(b) (6 points) *For* $x_i \in \{0,1\}$, *compute the following probability:*

$$\mathbb{P}\left(\text{The urn chosen is } i \mid X_1 = x_1, ..., X_n = x_n\right).$$

Show that this only depends on the number or red balls, i.e., $k = \sum_{i=1}^n x_i$.

Solution

(a) Let $k = \sum_{j=1}^n x_j$ the amount of red balls taken out. Then

$$\begin{aligned}\mathbb{P}\left(X_1 = 1, X_2 = 1\right) &= \mathbb{P}\left(X_1 = 1, X_2 = 1 \mid \text{Urn 1 is chosen}\right)\mathbb{P}\left(\text{Urn 1 is chosen}\right) \\ &\quad + \mathbb{P}\left(X_1 = 1, X_2 = 1 \mid \text{Urn 2 is chosen}\right)\mathbb{P}\left(\text{Urn 2 is chosen}\right) \\ &= \left(\frac{1}{3}\right)^2 \frac{1}{2} + \left(\frac{2}{3}\right)^2 \frac{1}{2} = \frac{5}{18}.\end{aligned}$$

We have that X_1 and X_2 are not independent because:

$$\begin{aligned}\mathbb{P}(\{X_1 = 1\}) &= \mathbb{P}(\{X_2 = 1\}) \\ &= \mathbb{P}\left(\{X_1 = 1\} \mid \text{Urn 1 is chosen}\right) + \mathbb{P}\left(\{X_1 = 1\} \mid \text{Urn 2 is chosen}\right) \\ &= \left(\frac{1}{3}\right)\frac{1}{2} + \left(\frac{2}{3}\right)\frac{1}{2} = \frac{1}{2}.\end{aligned}$$

And $\mathbb{P}(X_1 = 1, X_2 = 1) > \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1)$. This happens because the first variable gives "information" about which urn we have chosen so it also gives information about the second variable.

(b) Just by definition:

$$\begin{aligned}&\mathbb{P}\left(\text{The urn chosen is } i \mid X_1 = x_1, ..., X_n = x_n\right) \\ =&\frac{\mathbb{P}\left(\text{The urn chosen is } i, X_1 = x_1, ..., X_n = x_n\right)}{\mathbb{P}\left(X_1 = x_1, ..., X_n = x_n\right)} \\ =&\frac{\mathbb{P}\left(X_1 = x_1, ..., X_n = x_n \mid \text{Urn} = i\right)/2}{\sum_{j=1}^2 \mathbb{P}\left(X_1 = x_1, ..., X_n = x_n \mid \text{Urn} = j\right)/2} \\ =&\frac{i^k\left(3-i\right)^{n-k}}{2^{n-k} + 2^k}.\end{aligned}$$

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 $\mu_0 = 1$, 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

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

Determine the values of the constants $c_1 < c_2$ so that $\pi(1) = 0.02$ and the function $\pi(\mu)$ is *symmetric with respect around the point* $\mu = 1$.

Solution

Given that the mean of X_i is μ, $\overline{X}_n$ is normal with mean μ and variance $1/n$, hence $Y := \sqrt{n}(\overline{X}_n - \mu)$ is standard normal. The function Φ is the c.d.f. of Y.

$$\mathcal{R} = \{Y \notin [\sqrt{n}(c_1 - \mu), \sqrt{n}(c_2 - \mu)]\},$$

thus

$$\pi(\mu) = \Phi(\sqrt{n}(c_1 - \mu)) + 1 - \Phi(\sqrt{n}(c_2 - \mu)).$$

The function $\pi(\mu)$ is symmetric with respect around the point $\mu = 1$ implies that $c_2 - 1 = 1 - c_1$, which also implies that

$$1 - \Phi(\sqrt{n}(c_2 - 1)) = \Phi(\sqrt{n}(c_1 - 1)).$$

Since $\pi(1) = 0.02$, $\Phi(\sqrt{n}(c_1 - 1)) = 0.01$. We find in the formula sheet that $\Phi^{-1}(0.01) = -2.33$, hence $10(c_1 - 1) = -2.33$, namely $c_1 = 0.767$, and $c_2 = 1.233$.

10. (6 points) Let $(A_n)_{n\geq 1}$ be a sequence of events in a probability space. Suppose

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

Prove that

$$\mathbb{P}\big(\textstyle\bigcap_{k\in\mathbb{N}}\bigcup_{n\geq k} A_n\big) = 0\,.$$

(This is known as the first Borel-Cantelli lemma.)

Solution

Since $\sum\limits_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$,

$$\lim_{k\to\infty} \sum_{n=k}^{\infty} \mathbb{P}(A_n) = 0.$$

Thus, we have

$$\mathbb{P}(\bigcap_{k\in\mathbb{N}}\bigcup_{n\geq k} A_n) \leq \mathbb{P}(\bigcup_{n\geq k'} A_n) \leq \sum_{n=k'}^{\infty} \mathbb{P}(A_n), \quad \forall k' \in \mathbb{N},$$

which implies $\mathbb{P}(\bigcap_{k\in\mathbb{N}}\bigcup_{n\geq k} A_n) = 0$.

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 $i \neq j$),

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

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 $\mathbb{P}(B) > 0$.

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

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 $\cup_i B_i = \Omega$),

$$\mathbb{P}\left(A\right) = \sum_{i} \mathbb{P}\left(A \mid B_i\right)\, \mathbb{P}\left(B_i\right).$$

9. Bayes Formula. For a partition $\{B_i\}_{i=1}^d$, (see item 8), an event A and an integer $i \in \{1, \ldots, d\}$,
$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i)\ \mathbb{P}(B_i)}{\mathbb{P}(A)},$$
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,
$$F_X(x) := \mathbb{P}(X \le x) = \mathbb{P}(\{\omega \in \Omega\ :\ X(\omega) \le x\}), \quad x \in \mathbb{R}.$$
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 $X = (X_1, \dots, X_n)$ 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,

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