Statistics

TA Session 1

Hong Deng
h.deng@tinbergen.nl

Máté Kormos
m.kormos@tinbergen.nl

Tinbergen Institute

September 9, 2019

Overview

1. Introduction
   - Practical Matters
   - Exam Preparation
2. TA Exercises
   - Book 3rd edition/Ch4.50
   - Book 3rd edition/Ch2.59
   - Book 3rd edition/Ch3.48
   - Book 3rd edition/Ch4.36
   - (Book 3rd edition/Ch2.44)
   - (Book 3rd edition/Ch3.44)
   - (Book 3rd edition/Ch4.34)
3. Past Exam Questions
   - Exam Q1
   - Exam Q19
4. Reminder: HW Assignment

Overview

Practical Matters

TA sessions & Homework:

- 6 TA sessions
  - Mondays 11:00 a.m. (1H Presentation + 0.5H Q&A)
  - Some of the TA exercises with $(*)$ and past exam questions will be discussed (check https://staff.fnwi.uva.nl/p.j.c.spreij/onderwijs/TI/statistics.html often)
  - Schedule for the last TA session: Final Exam for year 2018 will be discussed
- 5 graded HW assignments + 1 optional HW assignment
  - Deadline: Fridays 12:00 noon
  - You may work in pairs (recommended)
  - Submission:
    - Format: LaTeX-typed document
      (Tip: *Overleaf* can be useful to create LaTeX files, see https://www.overleaf.com)
    - The last HW assignment will be counted towards your final score if you decide to submit it

Office Hour:

- Thursday 4:30-5:30 p.m. (TIA open area)

Exam Preparation & Study Advice

- Sketch for a study cycle:
  (1) Lecture on Wednesday
  $\Rightarrow$ Digest and practice old exam questions & TA exercises with asterisk
  $\longrightarrow$ (2) TA on next Monday
  $\Rightarrow$ Work on HW
  $\longrightarrow$ (3) HW Submission on next Friday (You will receive HW feedback two days later)
- Usually open book exam (Your own summary? Important! Textbook? Too heavy.)
- Practice with past exams in pairs/groups, as early as possible
- Know how different concepts are related ▸ related distributions

Overview

Book 3rd edition/Ch4.50

Question

Suppose that X_i, where $i = 1, 2, \ldots, n$, are independent random variables with $\mathbb{E}[X_i] = \mu$, and $\mathbb{V}[X_i] = \sigma^2$. Let $\bar{X} \equiv n^{-1} \sum_{i=1}^{n} X_i$. Show that $\mathbb{E}\left[\bar{X}\right] = \mu$, and $\mathbb{V}\left[\bar{X}\right] = \sigma^2/n$.

Book 3rd edition/Ch4.50

Question

Suppose that X_i, where $i = 1, 2, \ldots, n$, are independent random variables with $\mathbb{E}[X_i] = \mu$, and $\mathbb{V}[X_i] = \sigma^2$. Let $\bar{X} \equiv n^{-1} \sum_{i=1}^{n} X_i$. Show that $\mathbb{E}[\bar{X}] = \mu$, and $\mathbb{V}[\bar{X}] = \sigma^2/n$.

Solution Apply linearity of expectations:

$$\mathbb{E}[\bar{X}] = \mathbb{E}\left[n^{-1} \sum_{i=1}^{n} X_i\right] = n^{-1} \sum_{i=1}^{n} \mathbb{E}[X_i] = n^{-1} \sum_{i=1}^{n} \mu = \mu$$

$$\begin{aligned}
\mathbb{V}[\bar{x}] &= \mathbb{V}\left[n^{-1} \sum_{i=1}^{n} X_i\right] \\
&= n^{-2}\, \mathbb{V}\left[\sum_{i=1}^{n} X_i\right] \quad [\text{independence of } X_i\text{s}] \implies \\
&= n^{-2} \sum_{i=1}^{n} \mathbb{V}[X_i] = n^{-2} \sum_{i=1}^{n} \sigma^2 = \frac{\sigma^2}{n}
\end{aligned}$$

Book 3rd edition/Ch2.59

Question

Let $U \sim$Uniform$[-1, 1]$. Find the density function of $X \equiv U^2$.

Book 3rd edition/Ch2.59

Question

Let $U \sim$Uniform$[-1, 1]$. Find the density function of $X \equiv U^2$.

Solution First derive c.d.f.:

$$\begin{aligned}
F_X(x) &= \mathbb{P}(X \leqslant x) = \mathbb{P}\left(U^2 \leqslant x\right) \\
&= \mathbb{P}\left(-\sqrt{x} \leqslant U \leqslant \sqrt{x}\right) \\
&= \mathbb{P}\left(U \leqslant \sqrt{x}\right) - \mathbb{P}\left(U \leqslant -\sqrt{x}\right) \quad \text{[by property of probability measure]} \\
&= F_U(\sqrt{x}) - F_U(-\sqrt{x}) \quad \text{[by definition of } F\text{]} \\
&= \frac{\sqrt{x} - (-1)}{1 - (-1)} - \frac{-\sqrt{x} - (-1)}{1 - (-1)} = \sqrt{x} \implies \\
f_X(x) &= F_X'(x) = \frac{1}{2\sqrt{x}}
\end{aligned}$$

Book 3rd edition/Ch3.48

Question

Let T_1, T_2 independent exponentials with parameters λ_1, λ_2. Find the density of $T_1 + T_2$.

Book 3rd edition/Ch3.48

Question

Let T_1, T_2 independent exponentials with parameters λ_1, λ_2. Find the density of $T_1 + T_2$.

Solution Use the convolution rule. Let $Z \equiv T_1 + T_2$. Then for the joint density of T_1 and T_2, $f_{T_1,T_2}(t_1, t_2)$, we have

$$
\begin{aligned}
f_Z(z) &= \int_0^z f_{T_1,T_2}(x, z-x)dx \quad [\text{independence of } T_1,\ T_2] \implies \\
&= \int_0^z f_{T_1}(x) f_{T_2}(z-x)dx \quad [T_1,\ T_2 \text{ exponential}] \implies \\
&= \int_0^z \lambda_1 e^{-\lambda_1 x} \lambda_2 e^{-\lambda_2 (z-x)} dx = \lambda_1 \lambda_2 e^{-\lambda_2 z} \int_0^z e^{x(\lambda_2 - \lambda_1)} dx \\
&= \lambda_1 \lambda_2 e^{-\lambda_2 z} \left[\frac{e^{z(\lambda_2-\lambda_1)}}{\lambda_2 - \lambda_1} - \frac{1}{\lambda_2 - \lambda_1} \right] = \frac{\lambda_1 \lambda_2}{\lambda_2 - \lambda_1} \left[e^{-\lambda_1 z} - e^{-\lambda_2 z} \right]
\end{aligned}
$$

Book 3rd edition/Ch4.36 I

Question

Consider the following scheme for group testing. The original lot of samples (n) is divided into two subgroups, and each of the subgroups is tested as a whole. If either subgroup tests positive, it is divided into two, and the procedure is repeated. If any of the groups thus obtained tests positive, test every member of that group. Find the expected number of tests performed, and compare it to the number performed with no grouping and with the scheme described in Example C in Section 4.1.2. We further assume that (i) at each division step, the group is divided into two *equal* parts; and (ii) $n = 4K$ for some K strictly positive integer.

Solution See Appendix TA1.

Book 3rd edition/Ch4.36 II

Algorithm 1 Group Testing

input: $\mathcal{G}$: data set of n subjects
output: h: n-long vector $\{\ominus, \oplus\}^n$ of test results, $\ominus$ iff subject tests negative

$\mathcal{G}_A, \mathcal{G}_B \leftarrow$ split $\mathcal{G}$ in two equal, $n/2$ sized, groups randomly
for $j \in \{A, B\}$ **do**
    test $\mathcal{G}_j$
    if test is negative **then**
        $h_i \leftarrow \ominus \forall i \in \mathcal{G}_j$
    else
        $\mathcal{G}_{j1}, \mathcal{G}_{j2} \leftarrow$ split $\mathcal{G}_j$ in two equal, $n/4$ sized, groups randomly
        for $i \in \{1, 2\}$ **do**
            test $\mathcal{G}_{ji}$
            if test is negative **then**:
                $h_i \leftarrow \ominus \quad \forall i \in \mathcal{G}_{ji}$
            else
                test all subjects $\in \mathcal{G}_{ji}$ ($n/4$ tests): $h_i \leftarrow$test(i)$\in \{\ominus, \oplus\}$ for $i \in \mathcal{G}_{ji}$
return h

(Book 3rd edition/Ch2.44)

Question

Let T be an exponential random variable with parameter λ. Let X be a discrete random variable defined as $X = k$ if $k \leqslant T < k + 1$. $k = 0, 1, \ldots$ Find the frequency function for X.

(Book 3rd edition/Ch2.44)

Question

Let T be an exponential random variable with parameter λ. Let X be a discrete random variable defined as $X = k$ if $k \leqslant T < k+1$. $k = 0, 1, \ldots$ Find the frequency function for X.

Solution We will need cumulative distribution of T,

$$F_T(t) = \int_0^t f(s)ds = \int_0^t \lambda e^{-\lambda s} ds = 1 - e^{-\lambda t}$$

Note that $X = k$ if $T \in [k, k+1)$ for $k = 0, 1, \ldots$ Then

$$\begin{aligned} p(k) = P(X = k) =& P(k \leqslant T < k+1) \\ =& F_T(k+1) - F_T(k) \\ =& (1 - e^{-\lambda(k+1)}) - (1 - e^{-\lambda k}) \\ =& e^{-\lambda k} - e^{-\lambda(k+1)} \Rightarrow \\ p(k) =& e^{-\lambda k} \cdot (1 - e^{-\lambda}) \end{aligned}$$

Note: Geometric distribution $p(x) = (1-p)^x p$ where $p = (1 - e^{-\lambda})$ and $x = k$.

(Book 3rd edition/Ch3.44) I

Question

Let N_1 and N_2 be independent random variables following Poisson distributions with parameters λ_1 and λ_2, respectively. Show that the distribution of $N = N_1 + N_2$ is Poisson with parameter $\lambda = \lambda_1 + \lambda_2$.

Solution Let's call $N_1 = X$, $N_2 = Y$ and $N = Z$. Use **convolution**!

$$\begin{aligned} p_Z(z) &= \sum_{x=0}^{z} f_X(x) \cdot f_Y(z-x) \\ &= \sum_{x=0}^{z} \frac{\lambda_1^x}{x!} e^{-\lambda_1} \cdot \frac{\lambda_2^{z-x}}{(z-x)!} e^{-\lambda_2} \\ &= e^{-(\lambda_1+\lambda_2)} \cdot \sum_{x=0}^{z} \frac{\lambda_1^x \cdot \lambda_2^{z-x}}{x!(z-x)!} \end{aligned}$$

(Book 3rd edition/Ch3.44) II

Trick: Multiply and divide with $z!$ to use binomial coefficient $\frac{z!}{x!(z-x)!} = \binom{z}{x}$.

$$p_Z(z) = \frac{1}{z!} e^{-(\lambda_1+\lambda_2)} \cdot \sum_{x=0}^{z} \binom{z}{x} \lambda_1^x \cdot \lambda_2^{z-x}$$

Trick: Now we can use binomial formula $(a+b)^z = \sum_{x=0}^{z} \binom{z}{x} a^x b^{z-x}$

$$p_Z(z) = \frac{(\lambda_1+\lambda_2)^z}{z!} e^{-(\lambda_1+\lambda_2)}$$

Say $\lambda_1 + \lambda_2 = \lambda$, we get $p_Z(z) = \frac{\lambda^z}{z!} e^{-\lambda}$ which is again Poisson with parameter λ.

(Book 3rd edition/Ch4.34) I

Question

Let X be uniform on $[0, 1]$, and let $Y = \sqrt{X}$. Find $E(Y)$ by (a) finding the density of Y and then finding the expectation and (b) using Theorem A of Section 4.1.1.

Solution

(Method 1) Two-step procedure

Step 1: Density Transformation

(Book 3rd edition/Ch4.34) II

Either via cumulative distribution function...

$$\begin{aligned} F_Y(y) &= P(Y \leqslant y) \\ &= P(\sqrt{X} \leqslant y) \\ &= P(X \leqslant y^2) \\ &= y^2 \\ \Leftrightarrow f_Y(y) &= 2y \text{ for } y \in [0,1] \end{aligned}$$

(Book 3rd edition/Ch4.34) III

... or via (monotonic) density transformation function: use $f_Y(y) = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$.

$$\begin{aligned} f_X(x) &= 1,\ x \in [0,1] \\ y = g(x) = \sqrt{x} \Rightarrow x = g^{-1}(y) &= y^2 \\ \frac{dx}{dy} &= 2y \\ \Leftrightarrow \text{everything together: } f_Y(y) &= 2y \end{aligned}$$

Step 2: Expectation

$$E(Y) = \int_0^1 y f_Y(y)dy = \int_0^1 2y^2 dy = \frac{2}{3}y^3\Big|_0^1 = \frac{2}{3}$$

(Book 3rd edition/Ch4.34) IV

(Method 2) One-step procedure

Theorem (A in Section 4.1.1)

Suppose that $Y = g(X)$.

- *If X is discrete with $p(x)$, then $E(Y) = \sum_x g(x)p(x)$.*
- *If X is continuous with $f(x)$, then $E(Y) = \int_{-\infty}^{\infty} g(x)f(x)dx$.*[a]

[a] Provided that sum/integral is finite.

$\Rightarrow$ Using $y = g(x) = \sqrt{x}$ we get $E(Y) = \int_0^1 g(x)f_X(x)dx = \int_0^1 \sqrt{x}dx = \frac{2}{3}x^{\frac{3}{2}}\Big|_0^1 = \frac{2}{3}$.

Overview

Exam Q1

Question

Let U be a random variable that has uniform distribution on $[0,1]$. It is known that $\mathbb{E}[U] = \frac{1}{2}$ and $\mathbb{V}[U] = \frac{1}{12}$. Define other random variable X by $X = a + (b-a)U$ for $a < b$.
(a) Use the transformation rule for densities to show that X has a uniform distribution on $[a,b]$.
(b) Compute $\mathbb{E}[X]$ and $\mathbb{V}[X]$ from the definition of X.

Exam Q1

Question

Let U be a random variable that has uniform distribution on $[0, 1]$. It is known that $\mathbb{E}[U] = \frac{1}{2}$ and $\mathbb{V}[U] = \frac{1}{12}$. Define other random variable X by $X = a + (b - a)U$ for $a < b$.
(a) Use the transformation rule for densities to show that X has a uniform distribution on $[a, b]$.
(b) Compute $\mathbb{E}[X]$ and $\mathbb{V}[X]$ from the definition of X.

Solution
(a) Recall $f_X(x) = f_U(g^{-1}(x))\left|\frac{d}{du}g^{-1}(x)\right|$. Final result:

$$f_X(x) = \frac{1}{b-a} \quad (a < b) \text{ for } a \leqslant x \leqslant b \Rightarrow X \sim U[a, b]$$

(b) Recall linearity of expectation and properties of variance. Final result:

$$E(X) = \frac{a+b}{2} \quad Var(X) = \frac{(b-a)^2}{12}$$

Exam Q1

Question

Let U be a random variable that has uniform distribution on $[0,1]$. It is known that $\mathbb{E}[U] = \frac{1}{2}$ and $\mathbb{V}[U] = \frac{1}{12}$. Define other random variable Y by $Y = -\theta \log U$ for some $\theta > 0$.
(c) Compute for each $y > 0$ the probability $\mathbb{P}(Y > y)$. What is the density of Y? (d) If $f(u) = u \log u - u$ (for $u > 0$), then $f'(u) = \log u$. Use this to compute $\mathbb{E}[Y]$.

Exam Q1

Question

Let U be a random variable that has uniform distribution on $[0,1]$. It is known that $\mathbb{E}[U] = \frac{1}{2}$ and $\mathbb{V}[U] = \frac{1}{12}$. Define other random variable Y by $Y = -\theta \log U$ for some $\theta > 0$.
(c) Compute for each $y > 0$ the probability $\mathbb{P}(Y > y)$. What is the density of Y? (d) If $f(u) = u \log u - u$ (for $u > 0$), then $f'(u) = \log u$. Use this to compute $\mathbb{E}[Y]$.

Solution

(c) Recall the cdf of Exponential Distribution. Final result:

$$P(Y > y) = e^{\frac{-y}{\theta}}$$

$$f_Y(y) = F_Y'(y) = \frac{1}{\theta} e^{\frac{-y}{\theta}} \quad \text{Trick: } Y \sim Exp\left(\frac{1}{\theta}\right)$$

(d) Recall L'Hopital's Rule. Final result:

$$E(Y) = \theta$$

Exam Q19

Question

Let $\boldsymbol{X} = (X_1, X_2)^\top$ be a vector of independent random variables that both have a normal $N(0, \sigma^2)$ distribution ($\sigma^2 > 0$). Let $\boldsymbol{Y} = (Y_1, Y_2)^\top$ with $\boldsymbol{Y} = \boldsymbol{AX}$, where $\boldsymbol{A}$ is the matrix

$$\boldsymbol{A} = \begin{pmatrix} a & -1 \\ b & ab \end{pmatrix}$$

for real numbers a and $b (b \neq 0)$.
(a) Compute the covariance matrix of $\boldsymbol{Y}$.

Solution: (a) Hint: use proposition of covariance matrix $\boldsymbol{\Sigma_Y} = \boldsymbol{A\Sigma_X A}'$

$$\boldsymbol{\Sigma_Y} = \sigma^2 \begin{pmatrix} a^2+1 & 0 \\ 0 & b^2(a^2+1) \end{pmatrix}$$

Exam Q19

Question

Let $\boldsymbol{X} = (X_1, X_2)^\top$ be a vector of independent random variables that both have a normal $N(0, \sigma^2)$ distribution ($\sigma^2 > 0$). Let $\boldsymbol{Y} = (Y_1, Y_2)^\top$ with $\boldsymbol{Y} = \boldsymbol{AX}$, where $\boldsymbol{A}$ is the matrix

$$\boldsymbol{A} = \begin{pmatrix} a & -1 \\ b & ab \end{pmatrix}$$

for real numbers a and $b(b \neq 0)$.
(b) What is the distribution of $\boldsymbol{Y}$?
(c) Show that Y_1 and Y_2 are independent random variables.
(d) Show that Y_1^2 and Y_2^2 are independent random variables.

Solution: (b) Hint: special property of the bivariate normal distribution (see slides)
$\boldsymbol{Y} \sim MN_2(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y), \boldsymbol{\mu}_Y = (0,0)^\top, \boldsymbol{\Sigma}_Y$ as shown in question (a).
(c) Hint: independence $\Leftrightarrow$ uncorr. for bivariate normal distribution
(d) Hint: use $Y_1 \perp Y_2$ and definition of CDF

Exam Q19

Question

Let $\boldsymbol{X} = (X_1, X_2)^\top$ be a vector of independent random variables that both have a normal $N(0, \sigma^2)$ distribution $(\sigma^2 > 0)$. Let $\boldsymbol{Y} = (Y_1, Y_2)^\top$ with $\boldsymbol{Y} = \boldsymbol{AX}$, where $\boldsymbol{A}$ is the matrix

$$\boldsymbol{A} = \begin{pmatrix} a & -1 \\ b & ab \end{pmatrix}$$

for real numbers a and $b(b \neq 0)$.

(e) For certain real constants λ_1 and λ_2 put $U = \lambda_1 Y_1^2 + \lambda_2 Y_2^2$. How do we have to choose λ_1 and λ_2 such that U has a χ_2^2-distribution?

(f) How to choose the constants λ_1 and λ_2 in the previous part such that U has an exponential distribution with parameter 1?

Solution: (e) Hint: find a way to construct r.v. following standard normal distribution

$\lambda_1 = \frac{1}{a^2+1}, \lambda_2 = \frac{1}{b^2(a^2+1)}$

(f) Hint: $\chi_2^2 \implies \exp(\theta = 2)$ $\quad \lambda_1 = \frac{2}{a^2+1}, \lambda_2 = \frac{2}{b^2(a^2+1)}$

Overview

This Week's HW

Question

2nd Edition: Ch4.32, Ch4.45
3rd Edition: Ch4.32, Ch4.49

- **Deadline: 13 September, 12:00 noon**
- Please upload one copy of your *LaTeX-typed* solution on Canvas, with document title: "HW_n_<Your Name(s)>" ($n = 1, 2, \ldots, 6$)
- Better double check the conversion table, particularly for HW1 & 3 &5

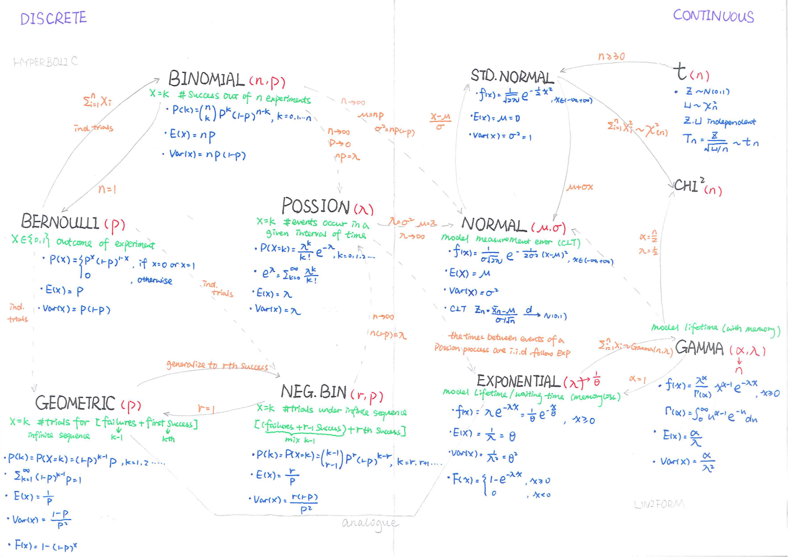

FIGURE: Wonderful Drawing of Distributions by previous TAs