

STAT 4100 Lecture Note

Week One (Sep 7 & 9, 2022)

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca, zhiyanggeezhou.github.io)

2022/Sep/09 15:51:28

IN THE CASE OF A FIRE ALARM:

- **Remain calm**
 - if it is safe, evacuate the classroom or lab
 - go to the closest fire exit
 - do not use the elevators
- **If you need assistance to evacuate the building, inform your professor or instructor immediately.**
- **If you need to report an incident or a person left behind during a building evacuation, report it to a fire warden or call security services 204-474-9341.**
 - **Do not** reenter the building until the “all clear” is declared by a fire warden, security services or the fire department.
- **Important: only those trained in the use of a fire extinguisher should attempt to operate one!**



Syllabus

Contact

- Zhiyang (Gee) Zhou
 - Email: zhiyang.zhou@umanitoba.ca
 - Homepage: zhiyanggeezhou.github.io
 - Office: 330 Machray Hall
- Ke (Kelly) Wang
 - Email: wangk9@myumanitoba.ca
 - Office: 355 Machray Hall

Timeline

- Lectures
 - Mon/Wed/Fri 13:30–14:20 at St. Paul's College 258
- Labs
 - Tue 16:30–18:00 at Buller 315
- Office Hours
 - (instructor) Mon 14:30–15:30 (or by appointment) at 330 Macray Hall
 - (TA) by appointment

Grading

- Assignment 20%
 - Submitted via Crowdmark
 - Attaching source codes if R is used in computation
 - Always including necessary interpretation
- Midterm 30%
 - In the week of Oct 17
 - Open-book, 3-hour, and online
- Final 50%
 - Open-book and in-person (?)

Materials

- References (recommended but NOT required)
 - (CB) Casella & Berger. 2002. *Statistical Inference*, 2nd Ed.
 - * 2 hardcopies reserved at the Jim Peebles Science and Technology Library
 - (HMC) Hogg, Mckean & Craig. 2018. *Introduction to Mathematical Statistics*, 8th Ed.
 - * Hardcopy of 6th Ed. available at the Jim Peebles Science and Technology Library
 - Salsburg. 2001. *The lady tasting tea: how statistics revolutionized science in the twentieth century*. New York: WH Freeman.
- Lecture notes
 - zhiyanggeezhou.github.io
 - UM Learn
 - **Subject to change** without prior notice
- Fall 2022 Syllabi Appendix

Outline

“All models are wrong, but some are useful.”

— George Box, *Journal of American Statistical Association* 1976

- What are statistical models?
 - Distributions of random variables (r.v.s) of interest
- Statistical inference
 - To answer questions on the underlying statistical models, e.g.,
 - * What is the model?
 - * Is the r.v. distributed as $\mathcal{N}(0, 1)$?
- Topics to be covered
 - Prerequisite
 - Estimation (finite/large sample, optimality)
 - Confidence interval (finite/large sample, interpretation)
 - Hypothesis testing (finite/large sample, optimality, interpretation)

Basics on random variables (CB/HMC Chp. 1)

Definitions

- Definition of r.v.: a real-valued function defined on a sample space Ω , i.e.,

$$X = X(\omega), \quad \omega \in \Omega$$

- Cumulative distribution function (cdf) of r.v. X

$$F_X(x) = \Pr(X \leq x)$$

- Right continuous
 - * Roughly speaking, a function is right-continuous if no jump occurs when the limit point is approached from the right
- Non-decreasing
- $F_X(-\infty) = 0$ and $F_X(\infty) = 1$

Example Lec1.1

- Given $p \in (0, 1)$, suppose

$$F(x) = \begin{cases} 1 - (1 - p)^{\lfloor x \rfloor}, & x \geq 1, \\ 0, & \text{otherwise,} \end{cases}$$

where $\lfloor x \rfloor$ represents the integer part of x . Show that F is a cdf.

- Hint: Check the right-continuity of F at positive integers.

Types of random variables

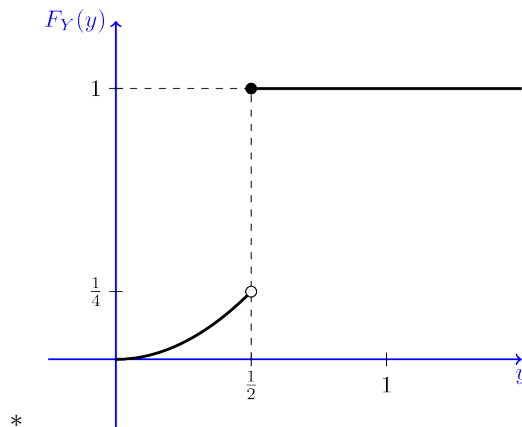
- X is a discrete r.v.
 - X takes countably many values
 - probability mass function (pmf): $p_X(x) = \Pr(X = x)$
- X is a continuous r.v.
 - cdf F_X is absolutely continuous, i.e., $\exists f_X$, s.t.

$$F_X(x) = \int_{-\infty}^x f_X(z) dz, \quad \forall x \in \mathbb{R}.$$

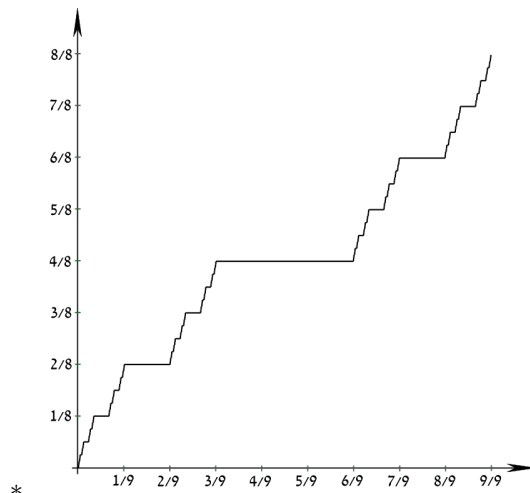
- f_X is the probability density function (pdf) of X
 - * $f_X(x) = (d/dx)F_X(x)$ if f_X is continuous at $x \in \mathbb{R}$

-
- Neither discrete nor continuous
 - X is a mixed r.v., e.g.,

$$F_X(x) = \begin{cases} 1, & x \geq 1/2; \\ x^2, & 0 \leq x < 1/2; \\ 0, & \text{otherwise.} \end{cases}$$



-
- Neither discrete nor continuous (con'd)
 - X is following the Cantor distribution



Univariate transformation (CB Sec. 2.1 & 2.4)

Support (CB pp. 50 & HMC pp. 46)

- In general, for real-valued function g
 - $\text{supp}(g) = \{x \in \text{domain}(g) : g(x) \neq 0\} \subset \text{domain}(g)$
- For discrete r.v. X
 - pmf $p_X(\cdot)$
 - $\text{supp}(X) = \text{supp}(p_X) = \{x \in \mathbb{R} : p_X(x) > 0\}$
 - e.g., support of $\text{Binom}(n, p)$ is $\{0, \dots, n\}$
- For continuous r.v. X
 - pdf $f_X(\cdot)$
 - $\text{supp}(X) = \text{supp}(f_X) = \{x \in \mathbb{R} : f_X(x) > 0\}$
 - e.g., support of $\mathcal{N}(0, 1)$ is \mathbb{R}

Indicator function

Given a set A , $\mathbf{1}_A(x) = 1$ if $x \in A$ and zero otherwise, i.e.,

$$\mathbf{1}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

Find pmf of $Y = g(X)$ given the pmf of X

1. Figure out $\text{supp}(Y) = \{y : y = g(x), x \in \text{supp}(X)\}$
2. Calculate $p_Y(y) = \Pr(Y = y) = \Pr(X \in \{x \in \text{supp}(X) : g(x) = y\})$

Example Lec2.1

Let X have the pmf $p_X(x) = 2^x \mathbf{1}_{\{-1, -2, \dots\}}(x)$. Find the pmf of $Y = X^4$.

Find cdf of $Y = g(X)$ given the distribution of X

- Calculate $F_Y(y) = \Pr\{g(X) \leq y\} = \Pr[X \in \{x : g(x) \leq y\}]$

Example Lec2.2

Let X have the uniform pdf $f_X(x) = \pi^{-1} \mathbf{1}_{(-\pi/2, \pi/2)}(x)$. Find the cdf of $Y = \tan X$.

cdf of $Y = F_X(X)$ (probability integral transformation, CB Thm. 2.1.10)

- If
 - $X \sim F_X$ (not necessarily continuous)
 - $Y = F_X(X)$
- Then $Y \sim \text{unif}(\text{image}(F_X))$
 - Specifically $Y \sim \text{unif}([0, 1])$ if X is continuous
- Application: inverse transform sampling
 - Goal: generate independent and identically distributed (iid) random samples following F_X
 - Implementation
 1. Sample iid $U_1, \dots, U_n \sim \text{unif}(\text{image}(F_X))$
 2. Then iid $F_X^{-1}(U_1), \dots, F_X^{-1}(U_n) \sim F_X$
 - * $F_X^{-1}(y) = \inf\{x : F_X(x) \geq y\}$
 - Pros & cons
 - * (Theoretically) applicable to arbitrary F_X
 - * The closed form of F_X^{-1} NOT always reachable

Find pdf of $Y = g(X)$ given the distribution of X

1. Figure out $\text{supp}(Y) = \{y : y = g(x), x \in \text{supp}(X)\}$
2. (Generically) If the cdf F_Y is known OR pdf f_X is easy to be integrated, then

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \int_{\{x: g(x) \leq y\}} f_X(x) dx$$

- The integration of f_X is often avoidable by employing the Leibniz Rule (CB Thm. 2.4.1):

$$\frac{d}{dy} \int_{a(y)}^{b(y)} f(x) dx = f\{b(y)\} \frac{d}{dy} b(y) - f\{a(y)\} \frac{d}{dy} a(y)$$

with $a(y)$ and $b(y)$ both differentiable with respect to y .

-
2. (Alternatively) According to CB Ex. 2.7(b), i.e., an extension of CB Thm. 2.1.5 & 2.1.8 and HMC Thm 1.7.1.

$$f_Y(y) = \sum_{k=1}^K f_X\{g_k^{-1}(y)\} \left| J_{g_k^{-1}} \right| \mathbf{1}_{B_k}(y)$$

- Partition $\text{supp}(X)$ into K intervals A_1, \dots, A_K such that $\bigcup_{k=1}^K A_k = \text{supp}(X)$ and $A_k \cap A_{k'} = \emptyset$ if $k \neq k'$
- g_k is strictly monotonic on A_k and $g(x) = g_k(x)$ for all $x \in A_k$
- g_k^{-1} is continuously differentiable on $B_k = \{g_k(x) : x \in A_k\}$
- Jacobian of transformation g_k^{-1}

$$J_{g_k^{-1}} = \frac{d}{dy} g_k^{-1}(y)$$

Example Lec2.2'

Let X have the uniform pdf $f_X(x) = \pi^{-1} \mathbf{1}_{(-\pi/2, \pi/2)}(x)$. Find the pdf of $Y = \tan X$.

Example Lec2.3

$X \sim \text{Weibull}(\text{shape} = \alpha, \text{scale} = \beta)$, viz. $f_X(x) = (\alpha/\beta)(x/\beta)^{\alpha-1} \exp\{-(x/\beta)^\alpha\} \mathbf{1}_{(0, \infty)}(x)$. Find the pdf of $Y = \ln(X)$.

Example Lec2.4

Let X have the pdf $f_X(x) = 2^{-1}\mathbf{1}_{(0,2)}(x)$. Find the pdf of $Y = X^2$.

Example Lec2.5

Let $f_X(x) = 3^{-1}\mathbf{1}_{(-1,2)}(x)$. Find the pdf of $Y = X^2$.