UNIVERSITY OF TORONTO | Statistical Sciences

DoSS Summer Bootcamp Probability Module 3

Ichiro Hashimoto

University of Toronto

July 15, 2024

Recap

Learnt in last module:

- Independence of events
  - Pairwise independence, mutual independence
  - Conditional independence
- Random variables
- Distribution functions
- Density functions and mass functions
- Independence of random variables

Outline

- Discrete probability
  - Classical probability
  - Combinatorics
  - Common discrete random variables
- Continuous probability
  - Geometric probability
  - Common continuous random variables
- Exponential family

Discrete probability

Example:

- Toss a fair coin, $P(H) = 1/2$ $= \mathbb{P}(T)$
- Roll a die, $P(\{1\}) = 1/6$

Discrete probability

Example:

- Toss a fair coin, $P(H) = 1/2$
- Roll a die, $P(\{1\}) = 1/6$

Classical probability

Classical probability is a simple form of probability that has equal odds of something happening.

Discrete probability

Example:

- Toss a fair coin, $P(H) = 1/2$
- Roll a die, $P(\{1\}) = 1/6$

Classical probability

Classical probability is a simple form of probability that has equal odds of something happening.

Remark:

For some event $A \in \mathcal{A}$, $\mathbb{P}(A)$ can be computed as the proportion:

$$\mathbb{P}(A) = \frac{\#\{\text{outcomes that satisfies } A\}}{\#\{\text{all the possible outcomes}\}}$$

Discrete probability

Converting the probability into counting problems

Permutations

For balls numbered 1 to n, choose r of them without replacement and record the order, the number of all the possible arrangements is

$$P(n, r) = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

1st ball 2nd ... r th ball

Remark:

Order matters.

$(1, 2)$ and $(2, 1)$ are considered different.

Discrete probability

Converting the probability into counting problems

Combinations

For balls numbered 1 to n, choose r of them without replacement regardless the order, the number of all the possible arrangements is

$$\binom{n}{r} = C_r^n = \frac{n!}{r!(n-r)!} = \frac{P(n,r)}{r!}$$

$r=3 \rightarrow$ A B C = CBA = BCA = ⋯

$3! = 6$ duplicates in permutation,

In general $r!$ duplicates in permutation.

Remark:

Order does not matter.

(1, 2) and (2, 1) are considered the same.

Discrete probability

Examples:

Forming committees: A school has 600 girls and 400 boys. A committee of 5 members is formed. What is the probability that it will contain exactly 4 girls?

Order does not matter

$$\frac{\overset{\text{4 girls}}{\binom{600}{4}} \times \overset{\text{1 boy}}{\binom{400}{1}}}{\binom{1000}{5}}$$

Discrete probability

Examples:

Forming committees: A school has 600 girls and 400 boys. A committee of 5 members is formed. What is the probability that it will contain exactly 4 girls?

Hypergeometric distribution

Randomly sample n objects without replacement from a source which contains a successes and $N - a$ failures, denote X as the number of successes. Then

N in total

$$\mathbb{P}(X = x) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}.$$

Discrete probability

Common discrete random variables

Bernoulli distribution

$\Omega = \{\text{failure, success}\}$, $X : \Omega \to \{0, 1\}$, and

$$\mathbb{P}(X = 1) = p, \quad \mathbb{P}(X = 0) = 1 - p.$$

Write $X \sim \mathit{Bernoulli}(p)$. $\iff \mathbb{P}(X=1) = p = 1 - \mathbb{P}(X=0)$

$\iff \mathbb{P}(X=x) = p^x (1-p)^{1-x}, \ x = 0, 1$

Discrete probability

Common discrete random variables

Bernoulli distribution

$\Omega = \{\text{failure, success}\}$, $X : \Omega \to \{0,1\}$, and

$$\mathbb{P}(X=1) = p, \quad \mathbb{P}(X=0) = 1-p.$$

Write $X \sim \mathit{Bernoulli}(p)$.

Example:

- Toss a coin $\text{Bern}\left(\frac{1}{2}\right)$
- Choose correct answer from A, B, C, D $\text{Bern}\left(\frac{1}{4}\right)$

Discrete probability

Common discrete random variables

Binomial distribution

Consider n independent Bernoulli trials with success probability $p \in (0,1)$, denote the number of successes as X. Then X can take values in $\{0, 1, \cdots, n\}$, and

$$\mathbb{P}(X = x) = \boxed{\binom{n}{x}} p^x (1-p)^{n-x}.$$

x successes

$n-x$ successes

Write $X \sim B(n, p)$.

Binom(n, p)

Order of successes and failure matter

Combination of x successes out of n trials $= \binom{n}{x}$

Discrete probability

$X = \sum_{i=1}^{n} Z_i$, $Z_i \overset{i.i.d}{\sim} Bern(p)$

i.i.d = independent and identically distributed

Common discrete random variables

Binomial distribution

Consider n independent Bernoulli trials with success probability $p \in (0,1)$, denote the number of successes as X. Then X can take values in $\{0, 1, \cdots, n\}$, and

$$\mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Write $X \sim B(n, p)$.

Example:

- Toss a coin 100 times
- Choose correct answer from A, B, C, D for 20 questions

UNIVERSITY OF TORONTO

Discrete probability

Common discrete random variables

Geometric distribution

Keep doing independent Bernoulli trials with success probability $p \in (0, 1)$ until the first success happens. Denote the number of trials as X. Then X can take values in $\{1, \cdots, \infty\}$, and

$$\mathbb{P}(X = x) = p(1-p)^{x-1} \leftarrow \text{geometric sequence w.r.t. } x$$

Write $X \sim Geo(p)$.

$$\underbrace{F \; - \; - \; - \; - \; F}_{x-1 \text{ failures}} \; \underset{x\text{th}}{S}$$

Discrete probability

Common discrete random variables

Geometric distribution

Keep doing independent Bernoulli trials with success probability $p \in (0,1)$ until the first success happens. Denote the number of trials as X. Then X can take values in $\{1, \cdots, \infty\}$, and

$$\mathbb{P}(X = x) = p(1-p)^{x-1}$$

Write $X \sim Geo(p)$.

Example:

- Toss a coin until the first head
- Choose answers from A, B, C, D until the first correct answer is picked

Discrete probability

Common discrete random variables

Negative binomial distribution

Keep doing independent Bernoulli trials with success probability $p \in (0,1)$ until the first r success happens. Denote the number of trials as X. Then X can take values in $\{r, \cdots, \infty\}$, and

$$\mathbb{P}(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}.$$

Write $X \sim \text{Neg-bin}(r, p)$.

$\underbrace{S \ F \ S \ S \ F \ F \ \text{-----} \ \text{---} \ F}_{x-1 \text{ trials}}$ S

rth success

$r-1$ successes

Discrete probability

Common discrete random variables

Negative binomial distribution

Keep doing independent Bernoulli trials with success probability $p \in (0,1)$ until the first r success happens. Denote the number of trials as X. Then X can take values in $\{r, \cdots, \infty\}$, and

$$\mathbb{P}(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}.$$

Write $X \sim \text{Neg-bin}(r, p)$.

Example:

- Toss a coin until the first 10 heads
- Choose answers from A, B, C, D until the first 3 correct answers are picked

Discrete probability

Common discrete random variables

Poisson distribution

Events occur in a fixed interval of time or space with a known constant mean rate λ and independently of the time since the last event, then denote the number of events during the fixed interval as X,

$$\mathbb{P}(X = x) = \frac{\lambda^x}{x!} exp(-\lambda).$$

Write $X \sim Poisson(\lambda)$.

decreases rapidly

$$\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = exp(\lambda)$$

taylor expansion,

Discrete probability

Common discrete random variables

Poisson distribution

Events occur in a fixed interval of time or space with a known constant mean rate λ and independently of the time since the last event, then denote the number of events during the fixed interval as X,

$$\mathbb{P}(X = x) = \frac{\lambda^x}{x!} exp(-\lambda).$$

Write $X \sim Poisson(\lambda)$.

Example:

fixed interval

- The number of patients arriving in an emergency room between 10 and 11 pm
- The number of laser photons hitting a detector in a particular time interval

UNIVERSITY OF TORONTO

Discrete probability

Common discrete random variables

Multinomial distribution

For n independent trials each of which leads to a success for exactly one of k categories, with each category having a given fixed success probability $p_i, i = 1, \cdots, k$, denote the number of successes of category i as X_i,

$$\mathbb{P}(X_1 = x_1, \cdots, X_k = x_k) = \binom{n}{x_1 x_2 \cdots x_k} p_1^{x_1} \cdots p_k^{x_k} \quad \text{with } \underline{\sum_{i=1}^{k} x_k = n}, \underline{\sum_{i=1}^{k} p_i = 1}.$$

Write $X \sim Multinomial(n, k, \{p_i\}_{i=1}^{k})$.

$$\binom{n}{x_1 x_2 \cdots x_k} \overset{def}{=} \frac{n!}{x_1! \, x_2! \cdots x_k!}$$

UNIVERSITY OF TORONTO

Out of $n!$ outcomes, we have the following duplicates:

C_1 : $x_1!$ duplicates

C_2 : $x_2!$ duplicates

$\vdots$

C_k : $x_k!$ duplicates

In total

$x_1!\, x_2! - - - - x_k!$ duplicates

Discrete probability

Common discrete random variables

Multinomial distribution

For n independent trials each of which leads to a success for exactly one of k categories, with each category having a given fixed success probability $p_i, i = 1, \cdots, k$, denote the number of successes of category i as X_i,

$$\mathbb{P}(X_1 = x_1, \cdots, X_k = x_k) = \binom{n}{x_1 x_2 \cdots x_k} p_1^{x_1} \cdots p_k^{x_k} \quad \text{with } \sum_{i=1}^{k} x_k = n, \sum_{i=1}^{k} p_i = 1.$$

Write $X \sim Multinomial(n, k, \{p_i\}_{i=1}^{k})$.

Remark:

The multinomial distribution gives the probability of any particular combination of numbers of successes for the various categories.

Continuous probability

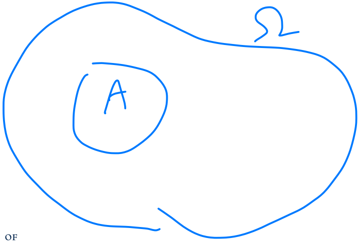
Example:

- Throw a needle evenly on a circle, the probability that it will lie in the left half.
- A bus arrives every 4 minutes, the probability of waiting for less than 1 minute.

Geometric probability

Geometric probability is a simple form of probability that has equal odds of something happening with infinite possible outcomes.

uniform distribution



$$\mathbb{P}(A) = \frac{\text{Area of } A}{\text{Area of } \Omega}.$$

Continuous probability

Example:

- Throw a needle evenly on a circle, the probability that it will lie in the left half.
- A bus arrives every 4 minutes, the probability of waiting for less than 1 minute.

Geometric probability

Geometric probability is a simple form of probability that has equal odds of something happening with infinite possible outcomes.

Remark:

For some event $A \in \mathcal{A}$, $\mathbb{P}(A)$ can be computed as the proportion:

$$\mathbb{P}(A) = \frac{\{\text{magnitude of outcomes that satisfies } A\}}{\{\text{magnitude of all the possible outcomes}\}}$$

Continuous probability

Common continuous random variables

(Continuous) Uniform distribution

X takes values in a fixed interval (a, b) evenly,

$$\begin{aligned} \mathbb{P}(X \leq x) &= \frac{x-a}{b-a}, \quad a \leq x \leq b, \\ f(x) &= \frac{1}{b-a}, \quad a \leq x \leq b. \end{aligned} \tag{1}$$

Write $X \sim U(a, b)$.

Unif (a,b)

Remark:

Continuous probability

Common continuous random variables

Normal distribution Gaussian

Define random variable X with the probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{(x-\mu)^2}{2\sigma^2}) \tag{2}$$

Write $X \sim N(\mu, \sigma^2)$.

mean variance

If $X \sim N(0, 1)$ standard normal

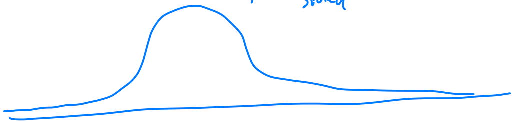
$f(x) = \frac{1}{\sqrt{2\pi}} exp\left(-\frac{x^2}{2}\right)$

Remark:

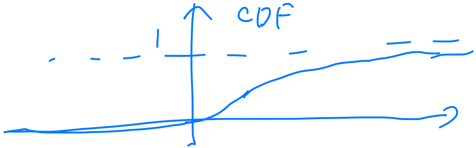
Most common distribution in nature

CLT.

two-sided



Continuous probability



Common continuous random variables

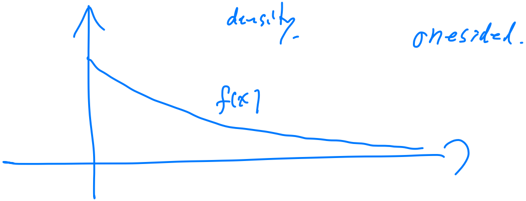
Exponential distribution

Define random variable X with the probability density function

$$\begin{aligned} P(X \leq x) &= 1 - \exp(-\lambda x), x \geq 0 \\ f(x) &= \lambda exp(-\lambda x), x \geq 0 \end{aligned} \tag{3}$$

Write $X \sim Exp(\lambda)$.

Remark:



Continuous probability

Common continuous random variables

Cauchy distribution

Define random variable X with the probability density function

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]} = \frac{1}{\pi\gamma}\left[\frac{\gamma^2}{(x-x_0)^2+\gamma^2}\right] \tag{4}$$

Write $X \sim Cauchy(x_0, \gamma)$. If $X \sim Cauchy(0, 1)$

$$f(x) = \frac{1}{\pi(1+x^2)}$$

Remark:

$$E|X| = 2\int_0^\infty \frac{x}{\pi(1+x^2)}\,dx = \frac{1}{\pi}\cdot \log(1+x^2)\Big|_0^\infty = \infty$$

Heavier tail!!

Continuous probability

Common continuous random variables

Gamma distribution

Define random variable X with the probability density function

one-sided.

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}e^{-\beta x}\beta^{\alpha}}{\Gamma(\alpha)} \quad \text{for } x>0 \quad \alpha,\beta>0, \tag{5}$$

normalising factor

Write $X \sim \Gamma(\alpha,\beta)$.

Remark:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-\beta x} \beta^\alpha \, dx$$

change variables $\beta x = u$.

$$= \int_0^\infty u^{\alpha-1} e^{-u} \, du \quad : \text{Gamma function.}$$

Properties of gamma function $\Gamma(\alpha)$

- $\Gamma(\alpha) = (\alpha-1)\,\Gamma(\alpha-1)$
- $\Gamma(1) = 1$
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

$\Rightarrow$ If $\alpha \in \mathbb{N}$,

$\Gamma(\alpha) = (\alpha-1)!$

Continuous probability

Common continuous random variables

Beta distribution

Define random variable X with the probability density function

$$f(x;\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \quad \text{for } 0 < x < 1 \quad \alpha,\beta > 0, \tag{6}$$

normalizing factor

bounded by both sides

Write $X \sim Beta(\alpha,\beta)$.

Remark:
$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx$. $= \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$

If $\alpha,\beta \in \mathbb{N}$, then $B(\alpha,\beta) = \frac{(\alpha-1)!\,(\beta-1)}{(\alpha+\beta-1)!}$

Exponential family

Consider a set of probability distributions whose pmf (discrete case) or pdf (continuous case) can be expressed in a certain form:

only depends on θ

Exponential family

parameter

$$f_X(x \mid \theta) = h(x) \exp\left[\eta(\theta) \cdot T(x) - A(\theta)\right], \tag{7}$$

where T, h are known functions of x; η, A are known functions of θ; θ is the parameter.

only depends on x

not on θ

Exponential family

Consider a set of probability distributions whose pmf (discrete case) or pdf (continuous case) can be expressed in a certain form:

Exponential family

$$f_X(x \mid \theta) = h(x) \exp\left[\eta(\theta) \cdot T(x) - A(\theta)\right], \tag{7}$$

where T, h are known functions of x; η, A are known functions of θ; θ is the parameter.

Merits:

- Facilitate the computation of some properties
- Bayesian statistics: conjugate prior
- Regression: GLM

Exponential family

Common distributions in the exponential family:

- Bernoulli / Binomial
- Poisson
- Negative Binomial
- Multinomial
- Exponential
- Normal
- Gamma
- Beta

check pmf

check density

UNIVERSITY OF TORONTO

Exponential family

Show that Bernoulli distribution belongs to the exponential family:

$X \sim \text{Bern}(p)$ $\qquad \mathbb{P}(X=1) = p = 1 - \mathbb{P}(X=0)$

$\iff \qquad \mathbb{P}(X=x) = p^x (1-p)^{1-x}$

$$\mathbb{P}(X=x) = p^x (1-p)^{1-x}$$

$$= \exp\left(\log\left\{ p^x (1-p)^{1-x} \right\} \right)$$

$$= \exp\left(x \log p + (1-x) \log(1-p) \right)$$

$$= \underbrace{1}_{h(x)} \cdot \exp\left(\underbrace{\boxed{\log \frac{p}{1-p}}}_{\eta(\theta)} \cdot \underbrace{x}_{T(x)} + \underbrace{\log(1-p)}_{-A(\theta)} \right)$$

Problem Set

Problem 1: The Robarts library has recently added a new printer which turns out to be defective. The letter "U" has a 30% chance of being printed out as "V", and the letter "V" has a 10% chance of being printed out as "U". Each letter is printed out independently, and all other letters are always correctly printed.
The librarian uses "UNIVERSITY OF TORONTO" as a test phrase, and will make a complaint to the printer factory immediately after the third incorrectly printed test phrase, calculate the probability that there are fewer correctly printed phrases than incorrectly printed phrases when the complaint is made.

Problem Set

Problem 2: Compute the mode of Negative binomial distribution with parameter r and p.
(Hint: consider $\mathbb{P}(X = k+1)/\mathbb{P}(X = k)$)

Problem 3: Show that normal distribution belongs to the exponential family.