UNIVERSITY OF TORONTO | Statistical Sciences

DoSS Summer Bootcamp Probability Module 4

Ichiro Hashimoto

University of Toronto

July 16, 2024

Recap

Learnt in last module:

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

Outline

- Joint and marginal distributions
  - ▷ Joint cumulative distribution function
  - ▷ Independence of continuous random variables
- Conditional distribution
- Functions of random variables
  - ▷ Convolutions
  - ▷ Change of variables
  - ▷ Order statistics

Joint and marginal distributions

Random vector: joint behaviour of multivariate random variables

Recall definition of random variable.

X is random variable. $\overset{\text{def.}}{\Longleftrightarrow}$ $\forall \underbrace{B \in \mathcal{R}}_{\text{Borel sets}}$, $X^{-1}(B) \in \mathcal{F}$

Instead of taking all Borel sets, we can check if X is random variable by special form of $B := (-\infty, x]$, $x \in \mathbb{R}$

$$X^{-1}((-\infty, x]) \in \mathcal{F} \text{ for any } x \in \mathbb{R}.$$

To define a random vector, we first generalize Borel sets to $\mathbb{R}^n$.

Let us consider the collection of cubes.

$$\{ (a_1, b_1] \times \cdots (a_n, b_n] \mid a_i < b_i, \; i = 1, \cdots, n \}.$$

Define $\mathcal{R}^n$ as the smallest σ-algebra containing all such cubes.

<u>Def</u> $X = (X_1, \cdots, X_n)$ is a random vector if

$$X^{-1}(B) \in \mathcal{F} \text{ for any } B \in \mathcal{R}^n.$$

<u>Remark.</u> Similarly with $n=1$, we can check if X is random vector by choosing special form of B,

$$B = (-\infty, x_1] \times \cdots \times (-\infty, x_n].$$

$$X^{-1}((-\infty, x_1] \times \cdots (-\infty, x_n]) \in \mathcal{F}, \text{ for any } x_1, \cdots, x_n \in \mathbb{R}$$

<u>Cor</u> If X_i $(1 \le i \le n)$ are random variable on $(\Omega, \mathcal{F})$, then $X = (X_1, \cdots, X_n)$ is a random vector.

(pf)

$$X^{-1}((-\infty, x_1] \times \cdots \times (-\infty, x_n]) = \{ X_i \le x_i, \; \forall i \}$$

$$= \bigcap_{i=1}^{n} \{ X_i \le x_i \}.$$

$$= \bigcap_{i=1}^{n} \underbrace{X_i^{-1}((-\infty, x_i])}_{\in \mathcal{F} \text{ since } X_i \text{ is random variable.}} \in \mathcal{F}.$$

intersection.

$$X_1 : \underbrace{\Omega_1}_{\substack{\Uparrow \\ \mathcal{F}_1}} \to \mathbb{R}, \qquad X_2 : \underbrace{\Omega_2}_{\substack{\Uparrow \\ \mathcal{F}_2}} \to \mathbb{R}.$$

$$X = (X_1, X_2) : \underbrace{\Omega_1 \times \Omega_2}_{\text{what is the natural way to define } \sigma\text{-algebra?}} \to \mathbb{R}^2.$$

Joint and marginal distributions

Random vector: joint behaviour of multivariate random variables

Joint cumulative distribution function

Consider a random vector $(X_1, X_2, \ldots, X_d)$, the joint cumulative distribution function of $(X_1, X_2, \ldots, X_d)$ is defined by:

$$F_{X_1,X_2,\cdots,X_d}(x_1, x_2, \cdots, x_d) = \mathbb{P}[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_d \leq x_d].$$

Joint and marginal distributions

Random vector: joint behaviour of multivariate random variables

Joint cumulative distribution function

Consider a random vector $(X_1, X_2, \ldots, X_d)$, the joint cumulative distribution function of $(X_1, X_2, \ldots, X_d)$ is defined by:

$$F_{X_1, X_2, \cdots, X_d}(x_1, x_2, \cdots, x_d) = \mathbb{P}[X_1 \leq x_1, X_2 \leq x_2, \ldots, X_d \leq x_d].$$

Remark:

For discrete random vector, it suffices to find the joint probability mass function

$$p_{X_1, \ldots, X_n}(x_1, \ldots, x_n) = \mathrm{P}(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n), \quad x_i \in \mathbb{R},$$

and

$$\mathbb{P}((X_1, \cdots, X_n) \in C) = \sum_{(x_1, \cdots, x_n) \in C} p_{X_1, \ldots, X_n}(x_1, \ldots, x_n).$$

UNIVERSITY OF TORONTO

Joint and marginal distributions

Remark:
For continuous random vector, consider the joint probability density function.

Joint probability density function

$$f_{X_1,\dots,X_n}(x_1,\dots,x_n) = \frac{\partial^n F_{X_1,\dots,X_n}(x_1,\dots,x_n)}{\partial x_1 \dots \partial x_n}, \quad x_i \in \mathbb{R}.$$

Similarly,

$$\mathbb{P}((X_1,\cdots,X_n) \in C) = \int_{(x_1,\cdots,x_n)\in C} f_{X_1,\dots,X_n}(x_1,\dots,x_n)\, dx_1 dx_2 \cdots dx_n.$$

Joint and marginal distributions

except X_i

Consider the special case of C where $X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n$ are allowed to take any possible values:

- Discrete case

$$\mathbb{P}(X_i = x_i) = \mathbb{P}(X_i = x_i, X_j \in \mathbb{R}, j \neq i) = \sum_{x_j, j \neq i} p_{X_1,\ldots,X_n}(x_1, \ldots, x_n).$$

- Continuous case

ith is missing.

$$\begin{aligned}\mathbb{P}(X_i \leq x_i) &= \mathbb{P}(X_i \leq x_i, X_j \in \mathbb{R}, j \neq i) \\ &= \int_{-\infty}^{x_i} \underbrace{\left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(t_1, \ldots, t_n)\, dt_1 \cdots dt_{i-1} dt_{i+1} \cdots dt_n \right)}_{\text{Integrate except for } t_i \to \text{ a function of } t_i} dt_i.\end{aligned}$$

Joint and marginal distributions

Taking the derivative regarding x_i, this gives us the marginal probability density function.

Marginal probability density function

$$f_{X_i}(x_i) = \underbrace{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1, \cdots, x_i, \cdots, x_n)\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n}_{\text{except for } x_i}.$$

Joint and marginal distributions

Taking the derivative regarding x_i, this gives us the marginal probability density function.

Marginal probability density function

$$f_{X_i}(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1, \cdots, x_i, \cdots, x_n)\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n.$$

Remark:
Marginal probability mass function (density function) of X_i is obtained by summing (integrating) the joint probability over all the other dimensions.

Joint and marginal distributions

Example: Draws from an urn

Suppose each of two urns contains twice as many red balls as blue balls, and no others, and suppose one ball is randomly selected from each urn, with the two draws independent of each other. Let A and B be discrete random variables associated with the outcomes of the draw from the first urn and second urn respectively. 1 represents a draw of red ball, while 0 represents a draw of blue ball.

joint

marginals

	1	0	$\mathbb{P}(B)$
1	$\frac{4}{9}$	$\frac{2}{9}$	$\frac{2}{3}$
0	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
$\mathbb{P}(A)$	$\frac{2}{3}$	$\frac{1}{3}$	1

Table: Joint and marginal pmf of draws from an urn

Joint and marginal distributions

Examples: continuous case

Consider the joint probability density function

$$f(x,y) = \begin{cases} kx & \text{for } 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$

Remark:

- Find k.
- Compute the marginal probability density function of X and Y.

Joint and marginal distributions

Integrate to find the value of k

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_0^1\int_0^1 kx\,dx\,dy = \int_0^1 \frac{k}{2}\,dy = \frac{k}{2}.$$

$$\therefore\ k = 2.$$

Marginal density

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^1 f(x,y)\,dy.$$

$$= \begin{cases} \int_0^1 2x\,dy & \text{--- if } x \in (0,1) \\ 0 & \text{--- if } x \notin (0,1) \end{cases}$$

UNIVERSITY OF TORONTO

$$= \begin{cases} 2x & \cdots \text{ if } x \in (0,1) \\ 0 & \cdots \text{ if } x \notin (0,1) \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^1 f(x,y)\,dx$$

$$= \begin{cases} \int_0^1 2x\,dx & \cdots \text{ if } y \in (0,1) \\ 0 & \cdots \text{ if } y \notin (0,1) \end{cases}$$

$$= \begin{cases} 1 & \cdots \text{ if } y \in (0,1) \\ 0 & \cdots \text{ if } y \notin (0,1) \end{cases}$$

Joint and marginal distributions

Recap: independence of random variables

Corollary of independence

If $X_1, \cdots, X_n$ are random variables, then $X_1, X_2, \cdots, X_n$ are independent if

$$P(X_1 \le x_1, \cdots, X_n \le x_n) = \prod_{i=1}^{n} P(X_i \le x_i)$$

Definition : $\mathbb{P}(X_1 \in A_1, \cdots, X_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i)$

for $\forall\, A_i \in \mathcal{R}$.

Joint and marginal distributions

Recap: independence of random variables

Corollary of independence

If $X_1, \cdots, X_n$ are random variables, then $X_1, X_2, \cdots, X_n$ are independent if

$$P(X_1 \le x_1, \cdots, X_n \le x_n) = \prod_{i=1}^{n} P(X_i \le x_i)$$

Remark:

Suppose $X_1, \cdots, X_n$ can only take values from $\{a_1, \cdots\}$, then X_i's are independent if

$$P(\cap\{X_i = a_i\}) = \prod_{i=1}^{n} P(X_i = a_i).$$

UNIVERSITY OF TORONTO

Joint and marginal distributions

Remark:
This is equivalent to check whether the joint pmf is always the product of the corresponding marginal pmf.

Generalize this to the continuous version:

Joint and marginal distributions

Remark:
This is equivalent to check whether the joint pmf is always the product of the corresponding marginal pmf.

Generalize this to the continuous version:

Independence of continuous random variables

Suppose $X_1, \cdots, X_n$ are continuous random variables, then X_i's are independent if

$$f_{(X_1,\cdots,X_n)}(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

Conditional distribution

Remark:

Given joint and marginal distributions, consider the conditional behaviour:

Conditional distribution

Remark:

Given joint and marginal distributions, consider the conditional behaviour:

Conditional distribution

For random variables X and Y, the conditional distribution of Y given $X = x$ is defined by

$$\frac{\mathbb{P}(Y=y \text{ and } X=x)}{\mathbb{P}(X=x)}$$

$$\|$$

- Discrete case

$$p_{Y|X=x}(y) = \mathbb{P}(Y = y \mid X = x) = \frac{p_{X,Y}(x, y)}{p_X(x)}.$$

- Continuous case

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

Conditional distribution

Remark:

Another look at independence:

- Discrete case:

  $$
  \begin{aligned}
  & X \text{ and } Y \text{ are independent} \\
  \Leftrightarrow\ & p_{Y|X=x}(y) = p_Y(y), \quad \forall x, y \\
  \Leftrightarrow\ & p_{X,Y}(x,y) = p_X(x)p_Y(y), \quad \forall x, y.
  \end{aligned}
  $$

- Continuous case:

  $$
  \begin{aligned}
  & X \text{ and } Y \text{ are independent} \\
  \Leftrightarrow\ & f_{Y|X}(y \mid x) = f_Y(y), \quad \forall x, y \\
  \Leftrightarrow\ & f_{X,Y}(x,y) = f_X(x)f_Y(y), \quad \forall x, y.
  \end{aligned}
  $$

Functions of random variables

Suppose we know the joint distribution of (X, Y), what is the distribution of $Z = X + Y$?

- Discrete case

$$\mathbb{P}(Z = z) = \sum_{x+y=z} \mathbb{P}(X = x, Y = y)$$

- Continuous case

$$\mathbb{P}(Z \leq z) = \int_{x+y\leq z} f_{X,Y}(x, y)\, dxdy$$

Remark:
This can be simplified in the independent case.

Functions of random variables

Convolutions of independent random variables

Suppose X and Y are independent, then for $Z = X + Y$,

- Discrete case

$$\mathbb{P}(Z = z) = \sum_{k=-\infty}^{\infty} \mathbb{P}(X = k)\mathbb{P}(Y = z - k).$$

- Continuous case

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z - x)\, dx$$

Sketch of proof: $\mathbb{P}(Z = z) = \sum_{x+y=z} \underbrace{\mathbb{P}(X = x, Y = y)}$

independence

$$= \sum_{x+y=z} \mathbb{P}(X = x)\,\mathbb{P}(Y = y)$$

$$= \sum_{x} \mathbb{P}(X = x)\,\mathbb{P}(Y = z - x)$$

Functions of random variables

Consider a function of random variable, and try to obtain the corresponding distribution function.

Multivariate change-of-variables formula

Suppose $\mathbf{X}$ is an n-dimensional random variable with joint density $f_{\mathbf{X}}(\mathbf{x})$. If $\mathbf{Y} = H(\mathbf{X})$, where H is a bijective, differentiable function, then $\mathbf{Y}$ has density $g_{\mathbf{Y}}(\mathbf{y})$:

$$g(\mathbf{y}) = f\left(H^{-1}(\mathbf{y})\right)\left|\det\left[\left.\frac{dH^{-1}(\mathbf{z})}{d\mathbf{z}}\right|_{\mathbf{z}=\mathbf{y}}\right]\right|$$

with the differential regarded as the Jacobian of $H^{-1}(\cdot)$, evaluated at $\mathbf{y}$.

Functions of random variables

Consider a function of random variable, and try to obtain the corresponding distribution function.

Multivariate change-of-variables formula

Suppose $\mathbf{X}$ is an n-dimensional random variable with joint density $f_{\mathbf{X}}(\mathbf{x})$. If $\mathbf{Y} = H(\mathbf{X})$, where H is a bijective, differentiable function, then $\mathbf{Y}$ has density $g_{\mathbf{Y}}(\mathbf{y})$:

$$g(\mathbf{y}) = f\left(H^{-1}(\mathbf{y})\right) \left| \det \left[\left. \frac{dH^{-1}(\mathbf{z})}{d\mathbf{z}} \right|_{\mathbf{z}=\mathbf{y}} \right] \right|$$

with the differential regarded as the Jacobian of $H(\cdot)$, evaluated at $\mathbf{y}$.

Remark:

Bijective property is important.

Functions of random variables

Special case of 2-dimensional vectors

2-dimensional change-of-variables formula

Suppose $\mathbf{X} = (X_1, X_2)$ with joint density $f_{X_1,X_2}(x_1, x_2)$. If $Y_1 = H_1(X_1, X_2)$, $Y_2 = H_2(X_1, X_2)$, where H is a bijective, differentiable function, then $\mathbf{Y} = (Y_1, Y_2)$ has density $g_{\mathbf{Y}}(y_1, y_2)$:

$$g(y_1, y_2) = f_{X_1,X_2}\left(H_1^{-1}(y_1, y_2), H_2^{-1}(y_1, y_2)\right) \underbrace{\left| \frac{\partial H_1^{-1}}{\partial y_1} \frac{\partial H_2^{-1}}{\partial y_2} - \frac{\partial H_1^{-1}}{\partial y_2} \frac{\partial H_2^{-1}}{\partial y_1} \right|.}_{\text{Jacobian.}}$$

Remark:

Functions of random variables

Remark:
Every continuous bijective function from $\mathbb{R}$ to $\mathbb{R}$ is strictly monotonic.

Functions of random variables

Remark:
Every continuous bijective function from $\mathbb{R}$ to $\mathbb{R}$ is strictly monotonic.

Special case of 1-dimensional random variable: generalize to monotonic functions

Univariate change-of-variables formula

Let $g : \mathbb{R} \to \mathbb{R}$ be a monotonic function on the support of $f_X(x)$, then for $Y = g(X)$, the density is:

$$f_Y(y) = f_X(g^{-1}(y)) \underbrace{\left| \frac{d}{dy}(g^{-1}(y)) \right|}_{\text{Jacobian.}} .$$

UNIVERSITY OF TORONTO

Functions of random variables

Proof of univariate change-of-variable formula:

Assume g : monotone, continuous, increasing

$$\int_{-\infty}^{y_0} f_Y(y)\, dy = \mathbb{P}(Y \leq y_0)$$

$$= \mathbb{P}(g(X) \leq y_0)$$

$$= \mathbb{P}(X \leq g^{-1}(y_0))$$

$$= \int_{-\infty}^{g^{-1}(z_0)} f_X(x)\,dx$$

$$\left[\begin{array}{l} \text{change of variable from calculus} \\ x = g^{-1}(z), \quad dx = \frac{d}{dz} g^{-1}(z) \end{array} \right]$$

$$= \int_{-\infty}^{z_0} \underbrace{f_X\left(g^{-1}(z)\right) \cdot \left| \frac{d}{dz} g^{-1}(z) \right|}_{\substack{\parallel \\ f_Y(z)}} dz$$

Hence desired formula follows.

$$\begin{tikzcd}
 & Y = g(X) \arrow[dl, "\text{induced}"'] & \\
\mu_Y \text{ on } \mathbb{R} \arrow[rr, leftrightarrow, bend right] & & \mu_X \text{ on } \mathbb{R} \arrow[ul, "\text{induced}"']
\end{tikzcd}$$

Relationship between these are given by change of variables formula.

Functions of random variables

Reordering index of X_i

Order statistics:

For random variables $X_1, X_2, \cdots, X_n$, the order statistics are $X_{(1)} \leq X_{(2)} \leq \cdots X_{(n)}$.

Cumulative distribution functions of order statistics

Consider the case where X_i's are independent identically distributed (i.i.d.) samples with cumulative distribution $F_X(x)$, then the CDF of $X_{(r)}$ satisfies

$$F_{X_{(r)}}(x) = \sum_{j=r}^{n} \binom{n}{j} [F_X(x)]^j [1 - F_X(x)]^{n-j},$$

the corresponding probability density function is

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!(n-r)!} f_X(x) [F_X(x)]^{r-1} [1 - F_X(x)]^{n-r}.$$

Functions of random variables

Special cases of $X_{(1)}$ and $X_{(n)}$:

$$F_{X_{(n)}}(x) = \mathbb{P}(\max\{X_1, \ldots, X_n\} \le x) = [F_X(x)]^n,$$

$$F_{X_{(1)}}(x) = \underbrace{\mathbb{P}(\min\{X_1, \ldots, X_n\} \le x)}_{= 1 - \mathbb{P}(\min\{x_1, \cdots, x_n\} > x)} = 1 - [1 - F_X(x)]^n.$$

Remark:

For continuous random variable, taking derivatives to obtain the probability density function.

$F_{X_{(r)}}(x) = \mathbb{P}(\underbrace{X_{(r)} \leq x})$

$\hookrightarrow$ rth smallest $X_i \leq x$

$\Longleftrightarrow$ there are "at least" r, X_i's $\leq x$.

$$= \sum_{j=r}^{n} \mathbb{P}(\text{there are exactly } j,\ X_i\text{'s} \leq x)$$

$$= \sum_{j=r}^{n} \underbrace{\binom{n}{r}}_{\text{combination of choosing } j \text{ out of } n.} F_X(x)^j \, (1 - F_X(x))^{n-j}$$

Problem Set

Problem 1: Show that the probability density function of normal distribution $N(\mu, \sigma^2)$ integrates to 1.
(Hint: consider two normal random variables X, Y)

Problem 2: Prove that for X with density function $f_X(x)$, the density function of $y = X^2$ is

$$f_Y(y) = \frac{1}{2\sqrt{y}}(f_X(-\sqrt{y}) + f_X(\sqrt{y})), \quad y \geq 0.$$

(Hint: start by considering the CDF)

Problem Set

Problem 3: Suppose $X_1, \cdots, X_n$ are i.i.d. sample following Uniform$[0, 1]$ distribution, find the joint probability density function of $(X_{(1)}, X_{(n)})$.
(Hint: start by considering the CDF)