UNIVERSITY OF TORONTO | Statistical Sciences

DoSS Summer Bootcamp Probability Module 4

Ichiro Hashimoto

University of Toronto

July 17, 2023

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 X is random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ if

$$X^{-1}(B) \in \mathcal{F} \quad \text{for any Borel set } B.$$

Instead of taking all Borel sets, we can check the definition by choosing special form of $B = (-\infty, x]$, $x \in \mathbb{R}$.

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

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

Let us consider the collection of cubes

$$\{ (a_1, b_1] \times \cdots \times (a_d, b_d] \mid a_i < b_i, \ i = 1, \cdots, d \}$$

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

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

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

<u>Remark</u> Similarly with $d=1$, we can check the definition by choosing special form of B,

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

So, checking $X^{-1}\left((-\infty, x_1] \times \cdots \times (-\infty, x_d]\right) \in \mathcal{F}$ for any $x_1, \cdots, x_d \in \mathbb{R}$

is enough to check if X is a random vector.

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

(Proof)

$$X^{-1}\left((-\infty, x_1] \times \cdots \times (-\infty, x_d]\right) = \{ X_i \le x_i, \ \forall i = 1, \cdots, d \}$$

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

$$= \underbrace{\bigcap_{i=1}^{d} \underbrace{X_i^{-1}\left((-\infty, x_i]\right)}_{\in \mathcal{F}}}_{\in \mathcal{F}} \in \mathcal{F} \quad //$$

Joint and marginal distributions

Random vector: joint behaviour of multivariate random variables

Joint cumulative distribution function

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

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

random variable, $F_X(x) = \mathbb{P}(X \le x)$

Joint and marginal distributions

Random vector: joint behaviour of multivariate random variables

Joint cumulative distribution function

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

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

Remark:

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

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

and

$$\mathbb{P}((X_1, \cdots, X_n) \in \underset{\substack{\cap \\ \mathbb{R}^n}}{C}) = \sum_{(x_1,\cdots,x_n)\in C} p_{X_1,\dots,X_n}(x_1, \dots, x_n).$$

Joint and marginal distributions

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

Joint probability density function

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

Similarly,

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

Joint and marginal distributions

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

- Discrete case

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

- Continuous case

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

Joint and marginal distributions

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

Marginal probability density function

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

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.

	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

marginals

UNIVERSITY OF TORONTO

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

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

$$\therefore k = 2.$$

Marginal density

$$f_X(x) = \int_0^1 f(x,y)\, dy = \int_0^1 2x\, dy = 2x \quad \text{on } 0 < x < 1$$

$$f_X = 0 \quad \text{otherwise.}$$

$$f_Z(z) = \int_0^1 f(x,z)\,dx = \int_0^1 2x\,dx = 1 \qquad \text{on } 0 < z < 1.$$

$$f_Z = 0 \qquad \text{otherwise.}$$

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

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

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

we can choose $A_i := (-\infty, x_i]$

UNIVERSITY OF TORONTO

Joint and marginal distributions

Recapture: 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

$$\underbrace{P(X_1 \leq x_1, \cdots, X_n \leq x_n)}_{\text{joint probability}} = \prod_{i=1}^{n} \underbrace{P(X_i \leq x_i)}_{\text{marginal probability}}$$

Remark:

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

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

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

- 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

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

$$\frac{f_{X,Y}(x,y)\Delta x \Delta y}{f_X(x)\Delta x} = \frac{f_{X,Y}(x,y)}{f_X(x)}\Delta y$$

Conditional distribution

Remark:

Another look at independence:

- Discrete case:

  X and Y are independent

  $\Leftrightarrow p_{Y|X=x}(y) = p_Y(y), \quad \forall x, y$

  $\Leftrightarrow p_{X,Y}(x,y) = p_X(x)p_Y(y), \quad \forall x, y.$

- Continuous case:

  X and Y are independent

  $\Leftrightarrow f_{Y|X}(y \mid x) = f_Y(y), \quad \forall x, y$

  $\Leftrightarrow f_{X,Y}(x,y) = f_X(x)f_Y(y), \quad \forall x, y.$

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: $\mathbb{P}(Z = z) = \sum_{x+y=z} \mathbb{P}(X = x, Y = y) = \sum_{k=-\infty}^{\infty} \mathbb{P}(X = k, \underbrace{Y = z - k}_{=z})$

independence

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

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 $g_{\mathbf{Y}}(\mathbf{y})$:

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

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

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 $g_{\mathbf{Y}}(\mathbf{y})$:

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

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:

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

"Real and complex Analysis" by Walter Rudin.

Functions of random variables

Proof of univariate change-of-variable formula:

$Y = g(X)$, g : monotone, increasing.

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

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

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

$$= \int_{-\infty}^{g^{-1}(y_0)} f_X(x)\, dx$$

$$\left[\, x = g^{-1}(y) \qquad dx = \frac{d}{dy} g^{-1}(y)\, dy \,\right]$$

$$= \int_{-\infty}^{y_0} \underbrace{f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|}_{\text{must be } = f_Y(y)} dy.$$

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

$Y = g(X)$

Y induces μ_Y on $\mathbb{R}$

X induce μ_X on $\mathbb{R}$

Relationship expressed through change of variables formula.

Functions of random variables

Rearranging indices 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

minimum of X_i

maximum of X_i

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

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

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

Remark:

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

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

rth smallest $X_i \leq x$

$\Longleftrightarrow$ there at least r X_i's $\leq x$

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

$$= \sum_{j=r}^{n} \underbrace{\binom{n}{j} F_X(x)^j (1 - F_X(x))^{n-j}}_{\text{exactly } j \ X_i\text{'s have } X_i \leq x.}$$

$$\frac{d}{dx} F_{X_{(r)}}(x) = \sum_{j=r}^{n} \binom{n}{j} \left\{ j f_X(x) F_X(x)^{j-1} (1 - F_X(x))^{n-j} - (n-j) f_X(x) F_X(x)^j (1 - F_X(x))^{n-j-1} \right\}$$

$= (0-0) + (0-0) + (0-0) \cdots$

$$= \underbrace{\binom{n}{r} r f_X(x) F_X(x)^{r-1} (1 - F_X(x))^{n-r}}_{\text{1st term of } r\text{th}}$$

$$+ \sum_{j=r}^{n-1} \left\{ \underbrace{\binom{n}{j+1}(j+1)}_{\text{1st term of } j+1\text{th}} - \underbrace{\binom{n}{j}(n-j)}_{\text{2nd term of } j\text{th.}} \right\} f_X(x) F_X(x)^j (1 - F(x))^{n-j-1}$$

$$+ \underbrace{0}_{\text{2nd term of } n\text{th}}$$

Note that $\binom{n}{j+1}(j+1) - \binom{n}{j}(n-j)$

$$= \frac{n!}{\cancel{(j+1)!}\,(n-j-1)!}\cancel{(j+1)} - \frac{n!}{j!\,\cancel{(n-j)!}}\cancel{(n-j)} = 0$$

$j!$ $\qquad$ $(n-j-1)!$

$$= \underbrace{\binom{n}{r} r}\; f_x(x)\, F(x)^{r-1}\left[1-F(x)\right]^{r}.$$

$$\left[\frac{n!}{r!\,(n-r)!}\, r = \frac{n!}{(r-1)!\,(n-r)!}\right]$$

$$= \frac{n!}{(r-1)!\,(n-r)!}\, f_x(x)\, F_x(x)^{r-1}\left[1-F_x(x)\right]^{r}.$$

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)