



UNIVERSITY OF
TORONTO

Statistical Sciences

DoSS Summer Bootcamp Probability Module 5

Ichiro Hashimoto

University of Toronto

July 15, 2026

Recap

Learnt in last module:

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

Outline

- Moments
 - ▷ Expectation, Raw moments, central moments
 - ▷ Moment-generating functions
- Change-of-variables using MGF
 - ▷ Gamma distribution
 - ▷ Chi square distribution
- Conditional expectation
 - ▷ Law of total expectation
 - ▷ Law of total variance

Measurable function

Measurable function

A map $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is called measurable if for any $B \in \mathcal{R}^m$ we have $f^{-1}(B) \in \mathcal{R}^n$. In other words, f is measurable if the preimage of any Borel set in $\mathbb{R}^m$ is also a Borel set in $\mathbb{R}^n$.

Remark: Recall that Borel σ -algebra contains almost all "nice" sets such as open sets, closed sets, and any sets that can be obtained by repeated use of countable union, intersections, and complements.

Function of random vector revisited

Measurable function of random vector is a random variable

Let $X = (X_1, \dots, X_n)$ is a random vector on $(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}^n$ and f be a measurable function on $\mathbb{R}^n$. Then $f(X)$ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow \mathbb{R}$.

Remark: The definition of a measurable function is given so that the above holds.

Proof:

Expectation

Intuition: How do the random variables behave on average?

Expectation

Intuition: How do the random variables behave on average?

Expectation

Consider a random variable X , the expectation of X is defined by $\mathbb{E}(X)$, where

- Discrete random variable

$$\mathbb{E}(X) = \sum_x x p_X(x),$$

- Continuous random variable

$$\mathbb{E}(X) = \int_{\mathbb{R}} x dF(x) = \int_{\mathbb{R}} x f_X(x) dx.$$

Expectation

Intuition: How do the random variables behave on average?

Measure theoretic definition of Expectation (informal)

Consider a random variable X , the expectation of X is defined by $\mathbb{E}(X)$, where

$$\mathbb{E}(X) = \lim_{n \rightarrow \infty} \sum_{i=-\infty}^{\infty} \frac{i}{n} \mathbb{P} \left(X \in \left(\frac{i}{n}, \frac{i+1}{n} \right] \right).$$

Remark: Recall that Borel σ -algebra and random variables are defined so that $\mathbb{P} \left(X \in \left(\frac{i}{n}, \frac{i+1}{n} \right] \right)$ is well defined.

Expectation

Expectation of function of random vector

Consider a random vector X and function $g(\cdot)$, the expectation of $g(X)$ is defined by $\mathbb{E}(g(X))$, where

- Discrete random vector

$$\mathbb{E}(g(X)) = \sum_x g(x) p_X(x),$$

- Continuous random vector in $\mathbb{R}^n$

$$\mathbb{E}(g(X)) = \int_{\mathbb{R}^n} g(x) dF(x) = \int_{\mathbb{R}^n} g(x) f_X(x) dx.$$

Expectation

Examples (random variable)

- $X \sim \text{Bernoulli}(p)$: $\mathbb{E}(X) = p \cdot 1 + (1 - p) \cdot 0 = p$.
- $X \sim \mathcal{N}(0, 1)$:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 0.$$

Expectation

Examples (random variable)

- $X \sim \text{Bernoulli}(p)$: $\mathbb{E}(X) = p \cdot 1 + (1 - p) \cdot 0 = p$.
- $X \sim \mathcal{N}(0, 1)$:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = 0.$$

Examples (random vector)

- $X_i \sim \text{Bernoulli}(p_i)$, $i = 1, 2$:

$$\mathbb{E}\left((X_1, X_2)^\top\right) = \left((\mathbb{E}(X_1), \mathbb{E}(X_2))^\top\right) = (p_1, p_2)^\top.$$

Expectation

Properties:

- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$;
- $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$;
- $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, when X, Y are independent.

Proof of the first property:

Moments

Raw moments

Consider a random variable X , the k -th (raw) moment of X is defined by $\mathbb{E}(X^k)$, where

- Discrete random variable

$$\mathbb{E}(X^k) = \sum_x x^k p_X(x),$$

- Continuous random variable

$$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} x^k dF(x) = \int_{-\infty}^{\infty} x^k f_X(x) dx.$$

Remark:

Moments

Central moments

Consider a random variable X , the k -th central moment of X is defined by $\mathbb{E}((X - \mathbb{E}(X))^k)$.

Remark:

- The first central moment is 0
- Variance is defined as the second central moment.

Variance

The variance of a random variable X is defined as

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Moments

Another look at the moments:

Moment generating function (1-dimensional)

For a random variable X , the moment generating function (MGF) is defined as

$$M_X(t) = \mathbb{E} \left[e^{tX} \right] = 1 + t\mathbb{E}(X) + \frac{t^2\mathbb{E}(X^2)}{2!} + \frac{t^3\mathbb{E}(X^3)}{3!} + \cdots + \frac{t^n\mathbb{E}(X^n)}{n!} + \cdots$$

Moments

Another look at the moments:

Moment generating function (1-dimensional)

For a random variable X , the moment generating function (MGF) is defined as

$$M_X(t) = \mathbb{E} \left[e^{tX} \right] = 1 + t\mathbb{E}(X) + \frac{t^2\mathbb{E}(X^2)}{2!} + \frac{t^3\mathbb{E}(X^3)}{3!} + \cdots + \frac{t^n\mathbb{E}(X^n)}{n!} + \cdots$$

Compute moments based on MGF:

Moments from MGF

$$\mathbb{E}(X^k) = \frac{d^k}{dt^k} M_X(t) \big|_{t=0}.$$

Moments

Relationship between MGF and probability distribution:

MGF uniquely defines the distribution of a random variable.

Moments

Relationship between MGF and probability distribution:

MGF uniquely defines the distribution of a random variable.

Example:

- $X \sim \text{Bernoulli}(p)$

$$M_X(t) = \mathbb{E}(e^{tX}) = e^0 \cdot (1 - p) + e^t \cdot p = pe^t + 1 - p.$$

- Conversely, if we know that

$$M_Y(t) = \frac{1}{3}e^t + \frac{2}{3},$$

it shows $Y \sim \text{Bernoulli}(p = \frac{1}{3})$.

Change-of-variables using MGF

Intuition: To get the distribution of a transformed random variable, it suffices to find its MGF first.

Properties:

- $Y = aX + b$, $M_Y(t) = \mathbb{E}(e^{t(aX+b)}) = e^{tb} M_X(at)$.
- $X_1, \dots, X_n$ independent, $Y = \sum_{i=1}^n X_i$, then $M_Y(t) = \prod_{i=1}^n M_{X_i}(t)$.

Change-of-variables using MGF

Intuition: To get the distribution of a transformed random variable, it suffices to find its MGF first.

Properties:

- $Y = aX + b$, $M_Y(t) = \mathbb{E}(e^{t(aX+b)}) = e^{tb} M_X(at)$.
- $X_1, \dots, X_n$ independent, $Y = \sum_{i=1}^n X_i$, then $M_Y(t) = \prod_{i=1}^n M_{X_i}(t)$.

Remark:

MGF is a useful tool to find the distribution of some transformed random variables, especially when

- The original random variable follows some special distribution, so that we already know / can compute the MGF.
- The transformation on the original variables is linear, say $\sum_i a_i X_i$.

Change-of-variables using MGF

Example: Gamma distribution

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

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

Compute the MGF of $X \sim \Gamma(\alpha, \beta)$ (details omitted),

$$M_X(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha} \quad \text{for } t < \beta, \text{ does not exist for } t \geq \beta.$$

Change-of-variables using MGF

Example: Gamma distribution

Observation:

The two parameters α, β play different roles in variable transformation.

- Summation:

If $X_i \sim \Gamma(\alpha_i, \beta)$, and X_i 's are independent, then $T = \sum_i X_i \sim \Gamma(\sum_i \alpha_i, \beta)$.

If $X_i \sim \text{Exp}(\lambda)$ (this is equivalently $\Gamma(\alpha_i = 1, \beta = \lambda)$) distribution, and X_i 's are independent, then $T = \sum_i X_i \sim \Gamma(n, \lambda)$.

- Scaling:

If $X \sim \Gamma(\alpha, \beta)$, then $Y = cX \sim \Gamma(\alpha, \frac{\beta}{c})$.

Change-of-variables using MGF

Example: χ^2 distribution

χ^2 distribution

If $X \sim \mathcal{N}(0, 1)$, then X^2 follows a $\chi^2(1)$ distribution.

Find the distribution of $\chi^2(1)$ distribution

- From PDF: (Module 4, Problem 2)

For X with density function $f_X(x)$, the density function of $Y = X^2$ is

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

this gives

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} \exp\left(-\frac{y}{2}\right).$$

Change-of-variables using MGF

Find the distribution of $\chi^2(1)$ distribution (continued)

- From MGF:

$$\begin{aligned}M_Y(t) &= \mathbb{E}(e^{tX^2}) = \int_{-\infty}^{\infty} \exp(tx^2) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \\&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2(1-2t)^{-1}}\right) dx \\&= (1-2t)^{-\frac{1}{2}} \int_{-\infty}^{\infty} \mathcal{N}(0, (1-2t)^{-1}) dx, \quad t < \frac{1}{2} \\&= (1-2t)^{-\frac{1}{2}}, \quad t < \frac{1}{2}.\end{aligned}$$

By observation, $\chi^2(1) = \Gamma(\frac{1}{2}, \frac{1}{2})$.

Change-of-variables using MGF

Generalize to the $\chi^2(d)$ distribution

$\chi^2(d)$ distribution

If $X_i, i = 1, \dots, d$ are i.i.d $\mathcal{N}(0, 1)$ random variables, then $\sum_{i=1}^d X_i^2 \sim \chi^2(d)$.

By properties of MGF, $\chi^2(d) = \Gamma(\frac{d}{2}, \frac{1}{2})$, and this gives the PDF of $\chi^2(d)$ distribution

$$\frac{x^{\frac{d}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{d}{2}} \Gamma(\frac{d}{2})} \quad \text{for } x > 0.$$

Conditional expectation

From expectation to conditional expectation:

How will the expectation change after conditioning on some information?

Conditional expectation

From expectation to conditional expectation:

How will the expectation change after conditioning on some information?

Conditional expectation

If X and Y are both discrete random vectors, then for function $g(\cdot)$,

- Discrete:

$$\mathbb{E}(g(X) \mid Y = y) = \sum_x g(x) p_{X|Y=y}(x) = \sum_x g(x) \frac{P(X = x, Y = y)}{P(Y = y)}$$

- Continuous:

$$\mathbb{E}(g(X) \mid Y = y) = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y) dx = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} g(x) f_{X,Y}(x, y) dx.$$

Conditional expectation

Properties:

- If X and Y are independent, then

$$\mathbb{E}(X \mid Y = y) = \mathbb{E}(X).$$

- If X is a function of Y , denote $X = g(Y)$, then

$$\mathbb{E}(X \mid Y = y) = g(y).$$

Sketch of proof:

Conditional expectation

Remark:

By changing the value of $Y = y$, $\mathbb{E}(X \mid Y = y)$ also changes, and $\mathbb{E}(X \mid Y)$ is a random variable (the randomness comes from Y).

Conditional expectation

Remark:

By changing the value of $Y = y$, $\mathbb{E}(X \mid Y = y)$ also changes, and $\mathbb{E}(X \mid Y)$ is a random variable (the randomness comes from Y).

Total expectation and conditional expectation

Law of total expectation

$$\mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}(X)$$

Proof: (discrete case)

Conditional expectation

Total variance and conditional variance

Conditional variance

$$\text{Var}(Y | X) = \mathbb{E}(Y^2 | X) - (\mathbb{E}(Y | X))^2.$$

Conditional expectation

Total variance and conditional variance

Conditional variance

$$\text{Var}(Y | X) = \mathbb{E}(Y^2 | X) - (\mathbb{E}(Y | X))^2.$$

Law of total variance

$$\text{Var}(Y) = \mathbb{E}[\text{Var}(Y | X)] + \text{Var}(\mathbb{E}[Y | X]).$$

Remark:

Problem Set

Problem 1: Prove that $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ when X and Y are independent.

(Hint: simply consider the continuous case, use the independent property of the joint pdf)

Problem 2: For $X \sim \text{Uniform}(a, b)$, compute $\mathbb{E}(X)$ and $\text{Var}(X)$.

Problem 3: Determine the MGF of $X \sim \mathcal{N}(\mu, \sigma^2)$.

(Hint: Start by considering the MGF of $Z \sim \mathcal{N}(0, 1)$, and then use the transformation $X = \mu + \sigma Z$)

Problem Set

Problem 4: The citizens of Remuera withdraw money from a cash machine according to $X = 50, 100, 200$ with probability $0.3, 0.5, 0.2$, respectively. The number of customers per day has the distribution $N \sim \text{Poisson}(\lambda = 10)$. Let $T_N = X_1 + X_2 + \cdots + X_N$ be the total amount of money withdrawn in a day, where each X_i has the probability above, and X_i 's are independent of each other and of N .

- Find $\mathbb{E}(T_N)$,
- Find $\text{Var}(T_N)$.