UNIVERSITY OF TORONTO | Statistical Sciences

DoSS Summer Bootcamp Probability Module 5

Ichiro Hashimoto

University of Toronto

July 19, 2023

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

Moments

Intuition: How do the random variables behave on average?

Moments

Intuition: How do the random variables behave on average?

Expectation

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

e.g. $g(x) = x_1 + \cdots + x_n$

where $x = (x_1, \cdots, x_n)^T$

- Discrete random vector

$$\mathbb{E}(g(X)) = \sum_{x = (x_1, \cdots, x_d)} g(x) p_X(x),$$

- Continuous random vector

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

$\mathbb{R}^d$ $\qquad$ $\mathbb{R}^d$

Recall the defn of EX is

$$EX = \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} \frac{k}{n} \mathbb{P}\left(X \in \left(\frac{k}{n}, \frac{k+1}{n}\right]\right)$$

We must have X be a random variable to make this definition valid.

i.e. $X^{-1}(B) \in \mathcal{F}$ for any Borel set $B \in \mathcal{R}$.

To make $E g(X)$ valid, we need $g(X)$ to be a random variable.

i.e. $\{g(X)\}^{-1}(B) \in \mathcal{F}$ for any Borel set $B \in \mathcal{R}$.

We can also write $\{g(X)\}^{-1}(B) = X^{-1}\left(g^{-1}(B)\right) \in \mathcal{F}$

$$\left[\; X^{-1}(g^{-1}(B)) \xrightarrow{X} g^{-1}(B) \xrightarrow{g} B \;\right]$$

Now, note that X is a random vector, Therefore,

if $g^{-1}(B) \in \mathcal{R}^d$ for any $B \in \mathcal{R}$, then

$X^{-1}\left(g^{-1}(B)\right) \in \mathcal{F}$ is ensured.

Def (measurable map / function)

A map (function) $f : (\Omega, \mathcal{F}) \to (\Xi, \hat{\mathcal{F}})$ is measurable

if $f^{-1}(A) \in \mathcal{F}$ for any $A \in \hat{\mathcal{F}}$.

Cor If $g : (\mathbb{R}^d, \mathcal{R}^d) \to (\mathbb{R}, \mathcal{R})$ is measurable, and

$X = (\Omega, \mathcal{F}) \to (\mathbb{R}^d, \mathcal{R}^d)$ is a random vector,

then $g(X)$ is a random variable.

$\longrightarrow$ We can define $E\, g(X)$

Problem What type of function $g : (\mathbb{R}^d, \mathcal{R}^d) \to (\mathbb{R}, \mathcal{R})$ is measurable?

1.) Indicator function $\mathbb{1}[x \in A]$ for $A \in \mathcal{R}^d$ is measurable.

$$\mathbb{1}[x \in A] = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

(Proof)

$$\mathbb{1}[x \in A]^{-1}(B) = \begin{cases} \phi & \text{if } 0 \notin B,\ 1 \notin B \\ A & \text{if } 1 \in B,\ 0 \notin B \\ A^c & \text{if } 0 \in B,\ 1 \notin B \\ \mathbb{R}^d & \text{otherwise.} \end{cases}$$

$\phi, \mathbb{R}^d \in \mathcal{R}^d$ trivially holds.

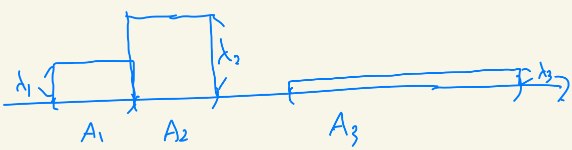
$A \in \mathcal{R}^d$ by assumption

$A^c \in \mathcal{R}^d$ since $\mathcal{R}^d$ is a σ-algebra.

Therefore, $\mathbb{1}[x \in A]^{-1}(B) \in \mathcal{R}^d$ holds always.

2) Simple function $g(x) = \sum_{k=1}^{n} \lambda_k \mathbb{1}[x \in A_k]$, $A_k \in \mathcal{R}^d$

simple functions are measurable.



3) limit of simple functions are measurable.

→ This includes all continuous functions and piecewise continuous functions.

→ Indeed almost any functions we use are measurable.

However, we can show the existence of non-measurable function.

Moments

Examples (random variable)

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

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \underbrace{x \frac{1}{\sqrt{2\pi}} exp(-\frac{x^2}{2})}_{\text{odd function}} \, dx = 0.$$

Moments

Examples (random variable)

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

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

Examples (random vector)

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

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

Moments

$\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$

$\mathbb{E}c = c$.
constant

$\mathbb{E}$ is linear.

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:

If discrete,

$$E(X+Y) = \sum_{k=-\infty}^{\infty} k \underbrace{P(X+Y=k)}_{\downarrow \text{ decompose.}}$$

$$= \sum_{k=-\infty}^{\infty} \underline{\sum_{j=-\infty}^{\infty}} k \underline{P(X=j, Y=k-j)}$$

Let $l = k-j \iff k = l+j$

$$= \sum_{l=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} (l+j) P(X=j, Y=l)$$

$$= \sum_{j=-\infty}^{\infty} j \underbrace{\sum_{l=-\infty}^{\infty} P(X=j, Y=l)}_{=P(X=j)} + \sum_{l=-\infty}^{\infty} l \underbrace{\sum_{j=-\infty}^{\infty} P(X=j, Y=l)}_{=P(Y=l)}$$

$$= \sum_{j=-\infty}^{\infty} j P(X=j) + \sum_{l=-\infty}^{\infty} l P(Y=l)$$

$$= EX + EY.$$

Q: How to prove this in general?

$$\mathbb{E}(X+Y) = \lim_{n\to\infty} \sum_{k=-\infty}^{\infty} \frac{k}{n} \underbrace{\mathbb{P}\left(X+Y \in \left(\frac{k}{n}, \frac{k+1}{n}\right]\right)}_{}$$

Can we decompose this probability in a similar manner?

$\longrightarrow$ We need quite sophisticated mathematical arguments to show even such a basic property.

Moments

higher-order moment

Raw moments

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

- Discrete random vector

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

- Continuous random vector

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

Remark:

Moments

Central moments

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

instead of X, use $X - \mathbb{E}X$

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

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

Moments

Another look at the moments:

Moment generating function (1-dimensional)

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

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

$t \in \mathbb{R}$

raw moments

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

kth (raw) moment.

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

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

Moments

Relationship between MGF and probability distribution:
MGF uniquely defines the distribution of a random variable.

If $M_X(t) = M_Y(t)$ on an open interval near 0,

then $X =_d Y$.

"X and Y have the same distribution"

Proof relies on Fourier Analysis

$\longrightarrow$ Billingsley "Probability"

Moments

Relationship between MGF and probability distribution:
MGF uniquely defines the distribution of a random variable.

Example:

- $X \sim Bernoulli(p)$

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

- Conversely, if we know that

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

Comparing these

$p = \frac{1}{3}$

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

by uniqueness of MGF.

Change-of-variables using MGF

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

$\mathbb{E}\, e^{tY}$ $\quad \mathbb{E}\, e^{tb} \cdot e^{taX} = e^{tb}\, \mathbb{E}\, e^{taX} = e^{tb} M_X(ta)$

Properties:

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

$$M_Y(t) = \mathbb{E} \exp(tY) = \mathbb{E} \exp\left(t \sum_i X_i\right) = \mathbb{E} \prod_{i=1}^n \exp(t X_i)$$

$$\text{by independence} = \prod_{i=1}^n \mathbb{E} \exp(t X_i) = \prod_{i=1}^n M_{X_i}(t)$$

Change-of-variables using MGF

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

Properties:

- $Y = aX + b$, $M_Y(t) = \mathbb{E}(e^{t(aX+b)}) = e^{tb}M_X(at)$.
- $X_1, \cdots, X_n$ 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),

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

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 $X_i \sim Exp(\lambda)$ (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 $Y = cX \sim \Gamma(\alpha, \frac{\beta}{c})$.

If $X_i \sim \Gamma(\alpha_i, \beta)$ and X_i's are independent,

$$M_{X_i}(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha_i}$$

So, for $Y = \sum_{i=1}^{n} X_i$,

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

$$= \prod_{i=1}^{n} \left(1 - \frac{t}{\beta}\right)^{-\alpha_i}$$

$$= \left(1 - \frac{t}{\beta}\right)^{-\sum_{i=1}^{n} \alpha_i}$$

$\xrightarrow{\text{By uniqueness of MGF}}$ $Y \sim \Gamma\left(\sum_i \alpha_i, \beta\right)$

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

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

$Y \sim \Gamma(\frac{1}{2}, \frac{1}{2})$

Change-of-variables using MGF

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

- From MGF:

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

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, $i = 1, \cdots, d$ 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:

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

UNIVERSITY OF TORONTO

Conditional expectation

Properties:

- If X and Y are independent, then

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

due to

$p_{X|Y}(x|y) = p_X(x)$

$f_{X|Y}(x|y) = f_X(x)$

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

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

Sketch of proof:

$$E[X \mid Y=y] = E[\underbrace{g(Y)}_{\text{constant given } Y=y} \mid Y=y] = g(y)$$

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).

$E(X|Y)$ can be treated as a function of Y

There is no randomness from X any more.

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)

$$LHS = \mathbb{E}\left[\sum_x x \cdot \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)} \right]$$

$$= \sum_y \left[\sum_x x \cdot \frac{\mathbb{P}(X=x, Y=y)}{\cancel{\mathbb{P}(Y=y)}} \right] \cancel{\mathbb{P}(Y=y)}$$

$$= \sum_x x \underbrace{\sum_z \mathbb{P}(X=x, Y=z)}_{= \mathbb{P}(X=x)} = \sum_x x \, \mathbb{P}(X=x) = \mathbb{E}X$$

(Continuous case.)

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}\left[\int_{-\infty}^{\infty} x f_{X,Y}(x,z)\,dx \Big/ f_Z(z) \right]$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x f_{X,Y}(x,z)\,dx \Big/ \cancel{f_Y(z)} \right] \cancel{f_Y(z)}\,dz$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \, f_{X,Y}(x,z)\,dx\,dz$$

$$= \int_{-\infty}^{\infty} x \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x,z)\,dz}_{= f_X(x)}\,dx$$

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

Conditional expectation

Total variance and conditional variance

Conditional variance

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

Conditional expectation

Total variance and conditional variance

Conditional variance

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

Law of total variance

$$Var(Y) = \underbrace{\mathbb{E}[\text{Var}(Y \mid X)]}_{\text{is random w.r.t. } X} + \text{Var}(\underbrace{\mathbb{E}[Y \mid X]}_{\text{is random w.r.t. } X}).$$

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 $X \sim \mathit{Uniform}(a, b)$, compute $\mathbb{E}(X)$ and $\mathrm{Var}(X)$.

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 $N \sim Poisson(\lambda = 10)$. Let $T_N = X_1 + X2 + \cdots + X_N$ 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 $Var(T_N)$.