UNIVERSITY OF TORONTO

Statistical Sciences

DoSS Summer Bootcamp Probability Module 5

Miaoshiqi (Shiki) Liu

University of Toronto

July 19, 2022

Recap

Learnt in last module:

- Joint and marginal distributions
  - ▷ Joint cumulative distribution function
  - ▷ Independence of continuous random variables
- Functions of random variables
  - ▷ Convolutions $Z = X + Y.$
  - ▷ 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?

$a_1 \quad a_2 \quad a_3 \quad \cdots \quad a_n$

min. max mean.

$X \qquad f_X(x).$

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

$a_1 \cdots a_n$ $\qquad \frac{\sum_{i=1}^{n} a_i}{n}$

- Discrete random vector

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

$$\sum_{i=1}^{n} \frac{1}{n} a_i$$

$$\boxed{\int_\Omega f \, d\mu}$$

- Continuous random vector

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

Moments

$X = \begin{cases} 1, & p \\ 0, & 1-p \end{cases}$

Examples (random variable)

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

$\uparrow \mu \quad \uparrow \sigma^2$

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

$X \sim N(\mu, \sigma^2)$. $\quad Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

$$E(X) = \int_{-\infty}^{+\infty} (\mu + \sigma z) \underbrace{\frac{1}{\sqrt{2\pi}} exp(-\frac{z^2}{2})}\, dz$$

$$= \mu.$$

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)

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

$$E\left(\begin{pmatrix} X_1 \\ X_2^2 \end{pmatrix}\right) = \begin{pmatrix} E(X_1) \\ E(X_2^2) \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$$

$$\mathbb{E}\left((X_1, X_2^2)^\top\right) = \left((\mathbb{E}(X_1), \mathbb{E}(X_2^2))^\top\right) = (p_1, p_2)^\top.$$

Moments

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

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:

$$X + Y = g\left(\begin{pmatrix} X \\ Y \end{pmatrix}\right)$$

$$E(X+Y) = \int_{\mathbb{R}^2} g\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) f_{X,Y}(x,y)\, dx\, dy.$$

$$= \int_{\mathbb{R}^2} (x+y) f_{X,Y}(x,y)\, dx\, dy.$$

$$= \int_{\mathbb{R}^2} x f_{X,Y}(x,y)\, dx\, dy + \int_{\mathbb{R}^2} y f_{X,Y}(x,y)\, dx\, dy.$$

$$\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} x f_{X,Y}(x,y)\, dy\right) dx$$

$$= \int_{-\infty}^{+\infty} x \left(\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dy\right) dx$$

$$= \int_{-\infty}^{+\infty} x f_X(x)\, dx$$

$$= E(X).$$

Moments

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: $k = 1$

$E(X^k) = E(X)$

Moments

Central moments

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

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

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

$$g(x) = e^{tX}$$

$$E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} f_X(x)\, dx < \infty.$$

UNIVERSITY OF TORONTO

Moments

$N(\mu, \sigma^2)$. $E(X) = \mu$.

$Var(x) = \sigma^2$

Another look at the moments:

Moment generating function (1-dimensional)

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

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

$k = 1$

$$\left.\frac{d}{dt} M_X(t)\right|_{t=0} = 0 + E(X) + \underbrace{t(\ldots)}$$

Compute moments based on MGF:

Moments from MGF

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

$$\frac{d}{dt} M_X(t) = E(X)$$

Moments

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

PDF / CDF / PMF $\longleftrightarrow$ MGF.

Moments

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

Example:

$$e^{tX} = \begin{cases} e^t, & p \\ 1, & 1-p. \end{cases}$$

- $X \sim Bernoulli(p)$

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

- Conversely, if we know that

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

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.

$M_X(t)$.

Properties:

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

$$M_Y(t) = E(e^{tY}) = E(e^{t(\sum_{i=1}^n X_i)}) = E(e^{tX_1 + tX_2 \cdots tX_n})$$

$$= E(\prod_{i=1}^n e^{tX_i})$$

$$= \prod_{i=1}^n \underbrace{E(e^{tX_i})}$$

UNIVERSITY OF TORONTO

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),

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

Change-of-variables using MGF

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

Observation:

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

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

- 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 $T = \sum_i X_i \sim \Gamma(n, \lambda)$.

  $\alpha = n.$

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

$= \left(1 - \frac{t}{(\beta/c)}\right)^{-\alpha}$

$= M_X(ct)$

$M_Y(t) = E(e^{tY}) = E(e^{ctX}) = E(e^{(ct)X})$

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

$Y = X^2$

- From PDF: (Module 4, Problem 2)
  For X with density function $f_X(x)$, the density function of $Y = X^2$ is

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

  this gives

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

UNIVERSITY OF TORONTO

Change-of-variables using MGF

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

- From MGF:

$$\exp\Big(\underbrace{tx^2 - \frac{x^2}{2}}\Big)$$

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

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

UNIVERSITY OF TORONTO

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

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

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

UNIVERSITY OF TORONTO

Conditional expectation

From expectation to conditional expectation:

How will the expectation change after conditioning on some information?

Conditional expectation

$g(X)$. $E(g(X)) =$ constant

$E(g(X) \mid Y) = f(Y)$.

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:

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

- Continuous:

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

UNIVERSITY OF TORONTO

Conditional expectation

Properties:

- If X and Y are independent, then

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

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

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

$X = g(Y)$

$Y = y \Rightarrow X = g(y).$

Sketch of proof:

discrete. X, Y. ind.

$$E(X \mid Y=y) = \sum_x x \underbrace{P_{X \mid Y=y}(x)} = \sum_x x P_X(x) = E(X).$$

$$E(X \mid Y=y) = \sum_x x \frac{P(X=x, Y=y)}{P(Y=y)} = \sum_{\{x=g(y)\}} x \frac{P(\cancel{Y=y})}{P(\cancel{Y=y})}$$

$$= g(y)$$

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).

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

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

$E(X)$ exists.

random about Y

$|E(X)| < \infty$.

Proof: (discrete case)

$$E(X \mid Y=y) = \sum_x x \, P_{X \mid Y=y}(x) = \sum_x x \frac{P(X=x, Y=y)}{P(Y=y)}$$

$$E(E(X \mid Y)) = \sum_y E(X \mid Y=y) P_Y(y)$$

$$= \sum_y \left(\sum_x x \frac{P(X=x, Y=y)}{P_Y(y)} \right) P_Y(y)$$

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

$$= \sum_x x \, P_X(x) = E(X)$$

Conditional expectation

Total variance and conditional variance

Conditional variance

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

$$Var(Y) = E(Y^2) - (E(Y))^2$$

UNIVERSITY OF TORONTO

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

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

$\{ Var(Y) < \infty. \quad E(Y^2) < \infty$

Remark: $Var(Y) = E(Y^2) - (E(Y))^2.$

$E(Var(Y|X)) = E(E(Y^2|X)) - E((E(Y|X))^2)$

$+ \qquad = E(Y^2) - E((\cancel{E(Y|X)})^2)$

$Var(E(Y|X)) = E((E(Y|X))^2) - (E(E(Y|X)))^2.$

$= E((\cancel{E(Y|X)})^2) - (E(Y))^2$

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 $Var(T_N)$.