UNIVERSITY OF TORONTO | Statistical Sciences

DoSS Summer Bootcamp Probability Module 5

Ichiro Hashimoto

University of Toronto

July 16, 2025

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

$\mathbb{E}X = \int x \, dF(x) = \int X \, d\mathbb{P} = \int x \, d\mathbb{P}_X(x)$

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

- Discrete random vector

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

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

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

Recall $\mathbb{E}X = \lim\limits_{n\to\infty} \sum\limits_{k=-\infty}^{\infty} \frac{k}{n} \mathbb{P}\Big(\underbrace{X \in \left(\frac{k}{n}, \frac{k+1}{n}\right]}\Big)$

$X^{-1}\left(\frac{k}{n}, \frac{k+1}{n}\right] \in \mathcal{F}$

$\mathbb{E}\,g(X)$ should be

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

when is $g(X)$ a random variable?

That is, when $g(X)^{-1}(B) \in \mathcal{R}^n$ for any $B \in \mathcal{R}$?

$$g(X)^{-1}(B) = X^{-1}\left(g^{-1}(B)\right)$$

$$\begin{array}{ccccc} \Omega & & \mathbb{R}^n & & \mathbb{R} \\ \cup & & \cup & & \cup \\ g(X)^{-1}(B) = X^{-1}(g^{-1}(B)) & \xrightarrow{X} & \boxed{g^{-1}(B)} & \xrightarrow{g} & \forall B \\ & & & & \in \\ & & & & \mathcal{R} \end{array}$$

needs to ensure $\in \mathcal{F}$

$g^{-1}(B)$ need also $\in \mathcal{R}^n$

<u>Def</u> $f: \mathbb{R}^n \to \mathbb{R}$ is called measurable if $\forall B \in \mathcal{R}$, $f^{-1}(B) \in \mathcal{R}^n$.

<u>Cor</u> If $f: (\mathbb{R}^n, \mathcal{R}^n) \to (\mathbb{R}, \mathcal{R})$ is measurable and $X: (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{R}^n)$ is a random vector, then $f(X): (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{R})$ is a random variable

Thus, if g is a measurable function,

$\mathbb{E}\, g(X)$ is a valid concept.

Q. What function is measurable?

1) Indicator function $\underbrace{\mathbb{1}_A(x)}$ for $A \in \mathcal{R}^n$ is measurable.

$$= \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

(proof)

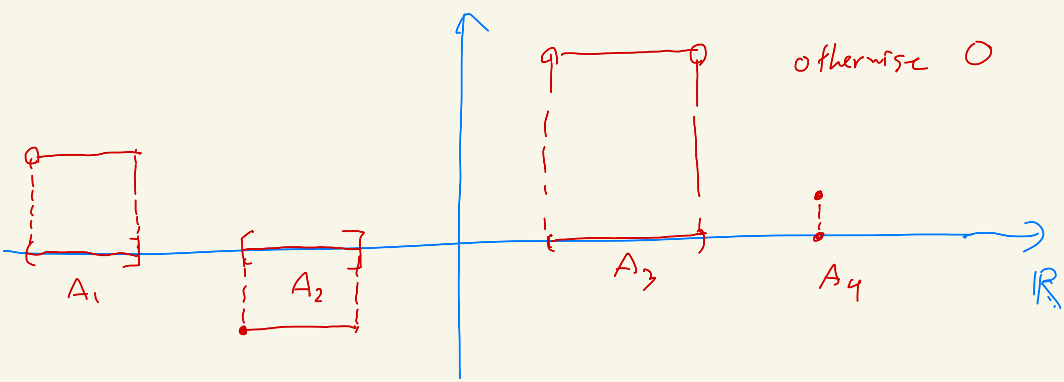
$$\mathbb{1}_A^{-1}(B) = \begin{cases} \phi \\ A \\ A^c \\ \mathbb{R}^n \end{cases}$$

there are only 4 possibilities and all of them belong to $\mathcal{R}^n$.

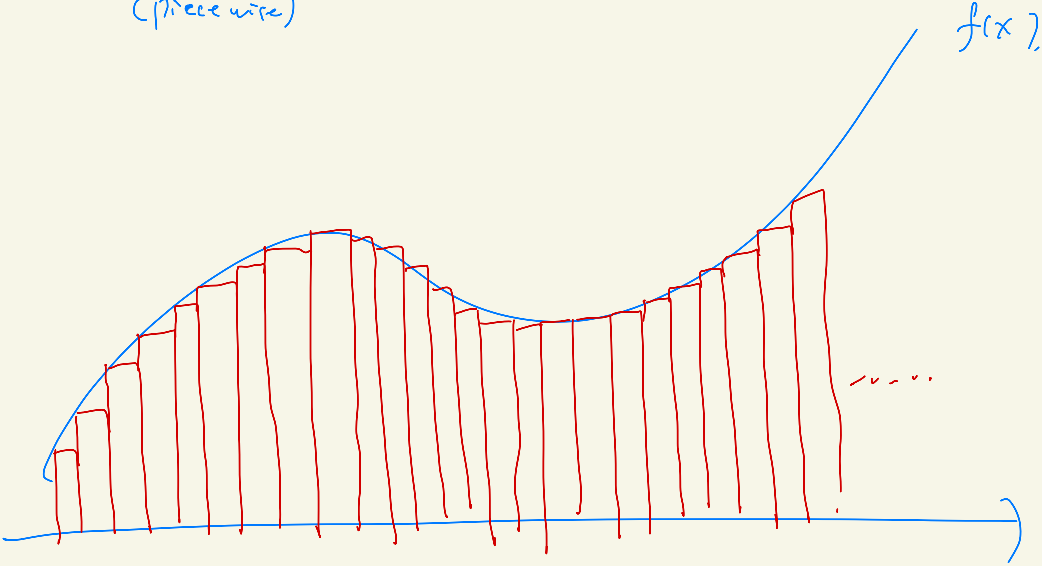
2) Simple functions

$$\underline{g(x) = \sum_{k=1}^{n} \lambda_k \mathbb{1}_{A_k}(x)}, \quad \lambda_k \in \mathbb{R}, \; A_k \in \mathcal{R}^n$$

linear combination of indicator functions.



choosing finer A_i's, you can approximate any continuous function by simple functions.
(piecewise)



3) Limits of simple functions are measurable.

$\longrightarrow$ all continuous functions are measurable.
(piecewise)

Moments

Examples (random variable)

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

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

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)

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

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

Moments

a real constant

$\mathbb{E}(b) = b$

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.

$\mathbb{E}$ is linear

$\mathbb{E}[aX+bY] = a\mathbb{E}X + b\mathbb{E}Y$.

b can be viewed as a random variable X s.t. $\mathbb{P}(X=b)=1$

$\mathbb{E}X = b \cdot \mathbb{P}(X=b) = b \cdot 1 = b$

Proof of the first property:

Assume X, Y are discrete and take integer values.

$$\mathbb{E}(X+Y) = \sum_{k=-\infty}^{\infty} k \underbrace{\mathbb{P}(X+Y=k)}_{=}$$

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

$k-j=l$.

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

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

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

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

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 $\mathbb{E}((X - \mathbb{E}(X))^k)$.

"centered"

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

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

a function of t

$$\left(\frac{d}{dt}\right)^k M_X(t)\Big|_{t=0} = \mathbb{E}X^k \quad k\text{-th moment}$$

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.

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

then $X =_d Y$

X and Y has same distribution

Proof relies on Fourier analysis

$\longrightarrow$ Billingley "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

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

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.

$= \mathbb{E}\, e^{atX} \cdot \underbrace{e^{tb}}_{\text{constant}} = e^{tb} \underbrace{\mathbb{E}\, e^{atX}}_{= M_X(at)}$

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

$M_Y(t) = \mathbb{E}\, e^{tY} = \mathbb{E}\, e^{t\sum_{i=1}^n X_i}$

$= \mathbb{E} \prod_{i=1}^n e^{tX_i}$ by independence

$= \prod_{i=1}^n \underbrace{\mathbb{E}\, e^{tX_i}}_{M_{X_i}(t)} = \prod_{i=1}^n M_{X_i}(t)$.

Change-of-variables using MGF

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

Properties:

- $Y = aX + b$, $M_Y(t) = \mathbb{E}(e^{t(aX+b)}) = e^{tb} M_X(at)$.
- $X_1, \cdots, X_n$ 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 $T = \sum_i X_i \sim \Gamma(\sum_i \alpha_i, \beta)$.
  special case. → 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 $Y = cX \sim \Gamma(\alpha, \frac{\beta}{c})$.

All of these can be shown by uniqueness of MGF.

UNIVERSITY OF TORONTO

If $X_i \sim \Gamma(\alpha_i, \beta)$, X_i's are indepent.

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

Since X_i's are indepent, for $T = \sum_{i=1}^{n} X_i$

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

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

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

By the uniqueness property of MGF,

$$T \sim \Gamma\left(\sum_{i=1}^{n} \alpha_i, \beta\right)$$

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

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

Change-of-variables using MGF

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) = 1$$

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

- From MGF:

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

set as σ^2

uniqueness of MGF.

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:

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

- Continuous:

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

Conditional expectation

Properties:

- If X and Y are independent, then

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

due to

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

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

if X and Y are independent

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

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

Sketch of proof:

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

$$\parallel$$

$$\mathbb{E}\left(g(Y) \mid Y = y \right)$$

$$\parallel$$

$$\mathbb{E}\left(\underbrace{g(y)}_{\text{constant}} \mid Y = y \right)$$

equal

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

$$\mathbb{E}(\underbrace{\mathbb{E}(X \mid Y)}_{\text{a function of } Y}) = \mathbb{E}(X)$$

expectation is over Y

Proof: (discrete case)

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

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

$$= \sum_y \sum_x x \, \mathbb{P}(X=x, Y=y)$$

(:)

$$= \sum_x x \underbrace{\sum_y \mathbb{P}(X=x, Y=y)}_{= \mathbb{P}(X=x)} = \sum_x x \, \mathbb{P}(X=x) = \mathbb{E}X.$$

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

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

Remark:

need this additional
term to recover Var(Y)

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 $X \sim \textit{Uniform}(a, b)$, 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 $N \sim Poisson(\lambda = 10)$. 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 $Var(T_N)$.