Lecture-03

03/03/2021

23/11/23

Review of probability, Covariance, Idependence

Note: The beginning Lecture on Lagrangian Optimization for Complex domain has been written as a continuation of Notes of Lecture-2.

- Random variable : $X \in \mathbb{R}$
- Probability Measure : $A \subseteq \mathbb{R}$ $P(A) = Pr\{x \in A\}$
- Conditional probability : $A, B \subseteq \mathbb{R}$ $P(A/B) = \frac{P(A \cap B)}{P(B)}$
- probability Density Function:

  $\hookrightarrow f_x : \mathbb{R} \rightarrow \mathbb{R}^+$

→ Random variable is something used to attach probabilities to Numbers. (The numbers are result of a Probability experiment)

→ The things which we deal are Probability Measures

probability are measures they behave like Volume, Mass ... and they are defined for every subset of Real Numbers

$$A \subseteq \mathbb{R} \qquad P(A) = Pr\{x \in A\}$$

Probability Measure is way of attaching measure to a set.

We know other kind of measures ex: Volume,

→ Conditional probability

$$\boxed{\text{Baye's Rule: } \quad P(A/B) = \frac{P(A \cap B)}{P(B)}}$$

→ Probability Density Function: $f_X(\xi)$

X → Random Variable

In this course it is a Fundamental Quantity

$$f_x : \mathbb{R} \longrightarrow R^+$$

It is defined in such a way that

$$P(A) = Pr\{x \in A\} = \int_A \underline{f_x(\xi)\,d\xi}$$

'A' is a set

Probability Density Function

$$f_X(\xi)\,d\xi \simeq Pr\left\{\xi - \frac{d\xi}{2} \leq X \leq \xi + \frac{d\xi}{2}\right\}$$

This approximation is true for small values of ξ

→ Cumulative Distribution function:

$$F_X(\; = Pr\{x \leq \xi\} = \boxed{\int_{-\infty}^{\xi} f_x(t)\,dt}$$

$$\boxed{f_X(\xi) = \frac{d\, F_X(\xi)}{d\xi}}$$

ξ is the value that the Random variable 'X' can take

→ Expectation:

$$\boxed{E[g(X)] = \int_{\mathbb{R}} g(\xi)\, f_X(\xi)\, d\xi}$$

This expectation of function of RV

weighted Average ~~w.r.t~~ wrto probabilities

$$\boxed{E[X] = \int_{\mathbb{R}} \xi\, f_X(\xi)\, d\xi}$$

Note: 'ξ' is the value that the RV 'X' takes

- Generalization:

$X \in \mathbb{R}^n$ → A Random vector of 'n' dimensions

$$f_X : \mathbb{R}^n \to \mathbb{R}^+$$

Therefore $P(A) = P_r\{x \in A\}$ $\quad A \subseteq \mathbb{R}^n$

$$P(A) = \int_A f_X(\xi)\, d\xi$$

→ here $\xi \in \mathbb{R}^n$

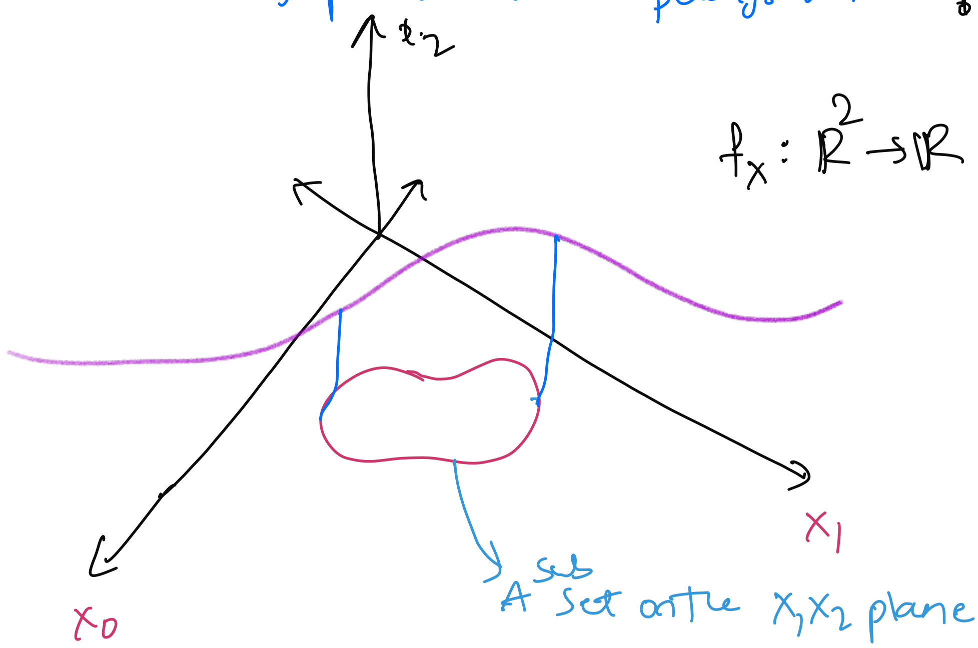
i.e. it is an n-dimensional

vector.

ex: $n = 2$ $X = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$

i.e every possible value belongs to the x_0-x_1 plane

$f_X : \mathbb{R}^2 \to \mathbb{R}$



Then $\boxed{P(A) = \int_A f_X(\xi) d\xi}$ → Here the 'X' RV is a 2-dimensional one
∴ 'ξ' is also 2D entity

This is similar to Multiple Random Variable

→ The Concept of Cumulative Distributive Functions can be extended to Multiple R.Vs

$$F_X(\xi) = \int_{-\infty}^{\xi_0} \cdots \int_{-\infty}^{\xi_{n-1}} f_X(t_0, \dots, t_{n-1})\, dt_0 \cdots dt_{n-1}$$

$X \in \mathbb{R}^n$

→

SPECIALIZATION of what we have discussed so far

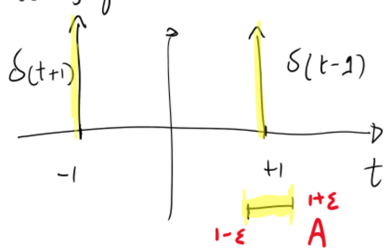
Dirac's delta is used to single out discrete values

head

$$\Pr\{c = +1\} = \frac{1}{2}$$

$$\Pr\{c = -1\} = \frac{1}{2}$$

tail



$$f_c(t) = \frac{1}{2}\delta(t-1) + \frac{1}{2}\delta(t+1)$$

$$\Pr\{x \in A\} = \int_{1-\varepsilon}^{1+\varepsilon} f_c(t)\,dt = \frac{1}{2}\underbrace{\int_{1-\varepsilon}^{1+\varepsilon} 1\,\delta(t-1)\,dt}_{1} + \frac{1}{2}\underbrace{\int_{1-\varepsilon}^{1+\varepsilon} \delta(t+1)\,dt}_{0} \rightarrow \int_{-\varepsilon}^{\varepsilon} \delta(t+2)\,dt$$

→ * The specialization is such that the numbers we want are random have generators in pathological in some way

This is the classical case which we would like to model

ex: Tossing of a coin

This is very common to assign to two different values

+1 → Head

-1 → Tail

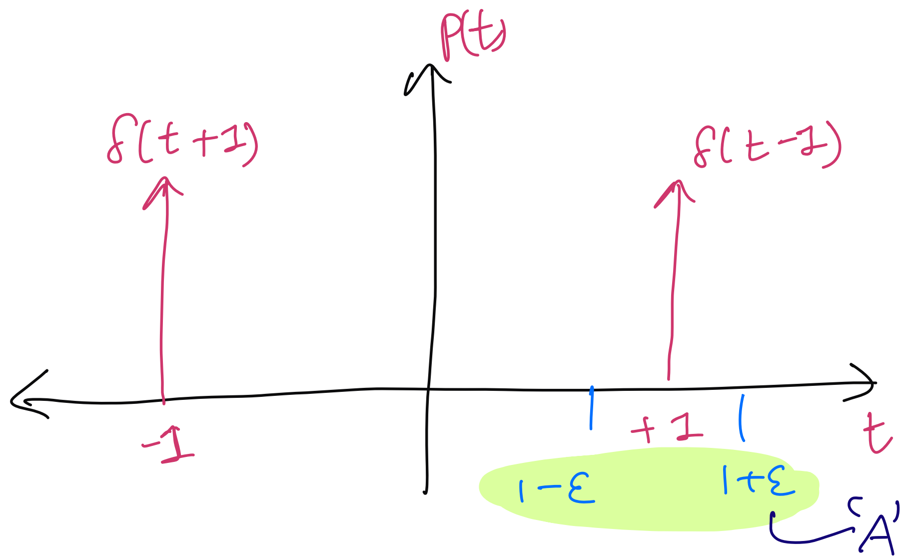
in Tossing a coin example

$$\text{Head} \rightarrow \Pr(C = \underset{}{\overset{\uparrow \text{Head}}{+1}}) = \tfrac{1}{2}$$

$$\text{Tail} \rightarrow \Pr(C = \underset{\downarrow \text{Tail}}{-1}) = \tfrac{1}{2}$$

Q) How do we Model This as Probability Distribution?

We have to resort to Dirac delta, when the no. of values we have is finite.



Basically, we are laying down a function that has zero all the time except t+1 & t-1 and we must also have the sum of two probability must be equal to 1.

∴ Integral of my PDF equal to '1'

$$\boxed{f_C(t) = \frac{1}{2}\,\delta(t-1) + \frac{1}{2}\,\delta(t+1)}$$

Ex: To calculate the Probability of 't' falling in the set A from $1-\varepsilon$ & $1+\varepsilon$

Probability is the Integral of the Probability Density function.

$$\int_{1-\varepsilon}^{1+\varepsilon} f_C(t)\,dt = \underbrace{\frac{1}{2}\int_{1-\varepsilon}^{1+\varepsilon}\delta(t-1)\,dt}_{1} + \underbrace{\frac{1}{2}\int_{1-\varepsilon}^{1+\varepsilon}\delta(t+1)\,dt}_{0}$$

In the domain

Not in the Domain

$$\Rightarrow \frac{1}{2}\cdot 1 = \frac{1}{2}$$

→ Expectation: For Random variable with a Prob. distribution. It's quite common to define a certain no. of scalars that these are numbers, not random variables, these are fixed.
These scalars are deterministic and characterize our Random variable.

2

BACK TO EXPECTATION

MOMENTS

$$m_x = \underset{\text{first order}}{E[x]} = \int_{-\infty}^{+\infty} \xi f_x(\xi)\, d\xi$$

$$\sigma_x^2 = \underset{\text{second order}}{E[(x-m_x)^2]} = \int_{-\infty}^{+\infty} (\xi - m_x)^2 f_x(\xi)\, d\xi$$

central moments of order p $\quad M_x^p = E[(x-m_x)^p]$

non-central moments of order p $\quad m_x^p = E[x^p]$

absolute moments of order p $\quad \mu_x^p = E[|x|^p]$

complex RV $z \in \mathbb{C} \implies \sigma_z^2 = E[|z-m_z|^2]$

modulus

Probably, the well know Expectation is the Average

Expectation:

$$\underset{1^{st}\text{ order}}{E[X]} = \int_{-\infty}^{\infty} \xi f_X(\xi)\, d\xi$$

Imp: 'ξ' is the Numerical value taken up by the our R.V 'X'

Variance:

2nd order

$$\sigma_X^2 = E[(X - m_X)^2]$$

$$\sigma_X^2 = \int_{-\infty}^{\infty} \underbrace{(\xi - m_X)^2 f_X(\xi)\, d\xi}$$

→ The above Expectation and Variance are the special cases of **Moments**.

1) $M_X^p = E[(X - m_X)^p]$ → Central Moment of order 'p'

2) $m_X^p = E[X^p]$ → Non-central Moment of order 'p'

$$= \int_{-\infty}^{\infty} \zeta^p f_X(\zeta)\, d\zeta$$

3) $\mu_X^p = E[|X|^p]$

4) For Complex RV

$$z \in \mathbb{C} \Rightarrow \sigma_z^2 = E\left[|z - m_z|^2\right]$$

modulus

→ Moments describe the information about the RV; all the information needed to describe a RV is in the Probability Density function.

Just with m_X & σ_X^2 we may know something about the behaviour of the Random Variable.

The best example about this could be the

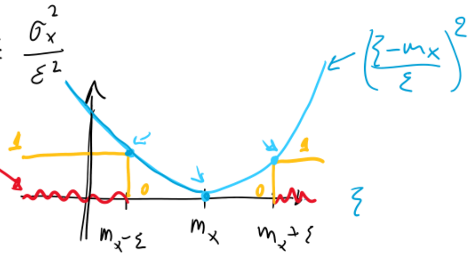
- Chebichev inequality.

$$\Pr\{\,|x-m_x|>\varepsilon\} \le \frac{\sigma_x^2}{\varepsilon^2}$$

Chebicev inequality

$x \in A =]-\infty, m_x-\varepsilon[\cup]m_x+\varepsilon, \infty]$

$$\Pr\{\underbrace{|x-m_x|>\varepsilon}\} \le \frac{\sigma_x^2}{\varepsilon^2}$$



Proof

$$\int_A f_x(\xi)\,d\xi = \int_{-\infty}^{+\infty} \chi_{m_x,\varepsilon}(\xi)\, f_x(\xi)\,d\xi$$

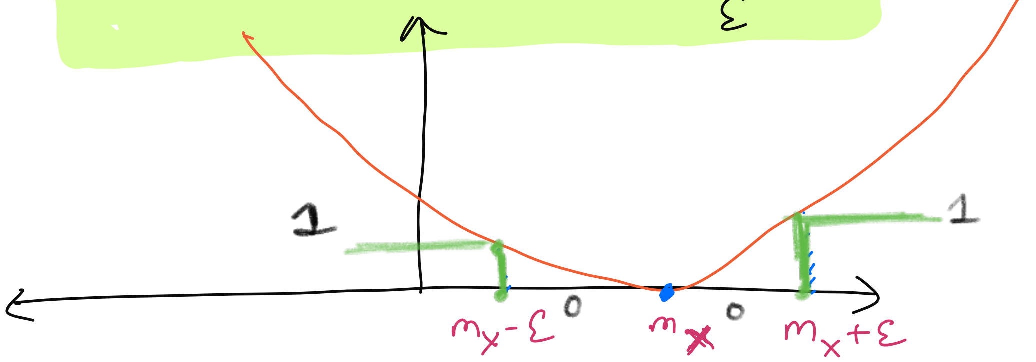
A $|x-m_x|>\varepsilon$

indicator function

$$\le \int_{-\infty}^{+\infty} \left(\frac{\xi-m_x}{\varepsilon}\right)^2 f_x(\xi)\,d\xi = \frac{1}{\varepsilon^2}\underbrace{\int_{-\infty}^{+\infty} (\xi-m_x)^2 f_x(\xi)\,d\xi} = \frac{\sigma_x^2}{\varepsilon^2}$$

It is important to understand the intuitive understand-ing of chebicev inequality.

$$\Pr\{|x-m_x|>\varepsilon\} \le \frac{\sigma_x^2}{\varepsilon^2}$$



→ $|x - m_x|$ is the distance of value of R.V 'x' Left/Right that has come out of a Random experiment.

→ The probability of $|x - m_x|$ is the probability of the Deviation.

So, $\underbrace{P_r\{|x - m_x| > \varepsilon\}}$ $= \dfrac{\sigma_x^2}{\varepsilon^2}$

① This is like asking what is the probability that value of R.V 'x' deviates from the Average m_x?

This Bounded by $\dfrac{\sigma^2}{\varepsilon^2}$

The larger the ε, the lower is the cap on the deviation. i.e probability of falling far from the average is always small.

proof of Chebichev inequality:

we will use our knowledge of calculating the Probability of a Set.

→ $$\int_{|x - m_x| > \varepsilon} f_x(\eta)\, d\eta$$

It is not comfortable to calculate such integral so we will use the trick of

↓
Domain of Integration

(Indicator function)

$$\Rightarrow \int_{|x-m_x|>\varepsilon} f_x(\eta)\,d\eta = \int_{-\infty}^{\infty} \chi_{m_x,\varepsilon}(\eta)\, f_x(\eta)\,d\eta$$

$$\chi_{m_x,\varepsilon}(\eta)$$

Indicator function.
(it is in Green in The Graph)

The indicator function of the set is a function that evaluates to 1. If the value is inside the set evaluates to 0 if the value is not inside the set.

→ The next set would be to calculate the Bounding

$$\int_{|x-m_x|>\varepsilon} f_x(\eta)\,d\eta \le \int_{-\infty}^{\infty} \left(\frac{\eta - m_x}{\varepsilon}\right)^2 f_x(\eta)\,d\eta$$

we are putting inside the Integer value larger than what I have to compute.

$\left(\frac{\eta - m_x}{\varepsilon}\right)^2$ is the parabola indicated above. graph
(It is in Red in the Graph)

$$\Rightarrow \int_{|x-m_x|>\varepsilon} f_x(\eta)\,d\eta \le \frac{1}{\varepsilon^2} \underbrace{\int_{-\infty}^{\infty} (\eta-m_x)^2 f_x(\eta)\,d\eta}_{=\sigma_x^2}$$

$$\Rightarrow \frac{1}{\varepsilon^2}\sigma_x^2$$

Covariance:-

→ Now, that we know the moments give us some information about RV we can extend the knowledge to a generalization of Variance.

It is very useful because it helps us to introduce the first processing task that we will see in the course.

It is Covariance, it is useful when the no. of RV involved is no longer one.

$$\gamma_{z_0 z_1} = E\left[(z_0 - m_{z_0})^* (z_1 - m_{z_1})\right]$$

for $z_0, z_1 \in \mathbb{C}$

$$z = \begin{bmatrix} z_0 \\ z_1 \end{bmatrix} \qquad f_z(\eta) : \mathbb{C}^2 \to \mathbb{R}^+$$

for $z_0 = z_1$ $\boxed{\gamma_{z_0 z_1} = E\left[|z_0 - m_{z_0}|^2\right]}$

This we have seen earlier

→ we think of two RV's z_1, z_2 as two components of a Random Vector!

i.e

$$z = \begin{bmatrix} z_0 \\ z_1 \end{bmatrix}$$

* (Complex Random Vector)

** * For this Random Vector we have a PDF attached and just to indicate it is the PDF of two things together it is called Joint PDF.

∴ The PDF of vector containing z_0 & z_1 is called Joint PDF of z_0 & z_1

∴ Exploiting this Joint PDF we can simply write the Expectation simply using the definition.

Joint PDF of z_0 & z_1

Covariance
$$\boxed{\gamma_{z_0 z_1} = \int_{\mathbb{C}}\int_{\mathbb{C}} (\zeta_0 - m_{z_0})^* (\zeta_1 - m_{z_1}) \underbrace{f_{\begin{bmatrix} z_0 \\ z_1 \end{bmatrix}}(\zeta_0, \zeta_1)}_{\text{Joint PDF}} d\zeta_0 \, d\zeta_1}$$

Q) How do you interpret the above equation?

This is the boiling down of the definition of

Covariance in terms of Integral and PDF

* Covariance is a strong indication that I can do or cannot do something which is called prediction.

It's a fact that the variables are RV but it doesn't mean that they don't have links with each other to certain Degree

& Establishing, Finding & Exploiting these LINKS is almost 50% of Statistical Signal processing.

→ Formally speaking when we want to use the Covariance is we want to predict Z_1 from value of Z_0

Ex: consider Z_0 & Z_1 are simply Temperatures but they are temperatures of two neighbouring rooms.

Temperature in a Room being a RV will have to measured to know, But since the Room are Close to each other the Temperature of one Room will probably affect Temperature of the neighbouring Room.

It's perfectly sensible to say let's measure the Temperature in One of them and let's predict the Temperature in the other.

predict z_1 from z_0

$$\hat{z}_1 = H(z_0)$$

prediction of z_1; it is not a true value it is just a value that I could measure for real.

The Idea is that we would like to make a predictor that makes a small error.

→ $$e = E\left[\,|z_1 - \hat{z}_1|^2\,\right] \Rightarrow$$ Average quadratic error

(Error)

is the function of z_0 which is a RV $\therefore$ $\hat{z}_1$ is also a RV

* → When it comes to predictors from now we will restrict ourselves to Linear predictors or Affine predictors

$a, b \in \mathbb{C}$

$$\hat{z}_1 = H(z_0) = a z_0 + b$$

Affine predictor

Predictors parameters

→ The Design of a Predictor task can be thought of as an optimization problem.

Because, what we actually want is this

$$\min_{a,b} E\left[|z_1-\hat{z}_1|^2\right] =$$

s.t (without any constraints)

a, b are our degrees of freedom

$$E\left[|z_1 - az_0 - b|^2\right] = L(a,b)$$

$$\Rightarrow E\left[(z_1-az_0-b)^*(z_1-az_0-b)\right] - \sum_{j=0}^{m-1} \lambda_1 E_j$$

(Since we don't have any constraints) $=0$

$$\Rightarrow E\left[(z_1-az_0-b)^*(z_1-az_0-b)\right] = 0$$

Here we can't make any Complex derivation because we have z^*

→ But Now we can follow Brandwood's Recipe which we studied earlier

Highlight the conjugate

$$E\left[(z_1-az_0-b)^*(z_1-az_0-b)\right] = L(a,b)=0$$

$$(z_1 - a z_0 - b)^* = z_1^* - a^* z_0^* - b^*$$

$$\tilde{\nabla}_{a^*, b^*} L(a,b) = 0$$

Here a, b are fixed

we can do the derivative w.r.t. non conjugate variables also

i.e $\tilde{\nabla}_{a,b} L(a,b) = 0$

Here a^*, b^* are fixed

$$\frac{\tilde{\partial} L}{\partial a^*} = E\left[z_0^* (z_1 - a z_0 - b) \right] = 0$$

$$\frac{\tilde{\partial}}{\partial a^*} E\left[(z_1^* - a^* z_0^* - b^*) \underbrace{(z_1 - a z_0 - b)}_{\text{constant in this case}} \right] = 0$$

$$\Rightarrow E\left[z_0^* (z_1 - a z_0 - b) \right] = 0 \quad \text{(1)}$$

$$\frac{\tilde{\partial} L}{\partial b^*} = E\left[-1 (z_1 - a z_0 - b) \right] = 0 \quad \text{(2)}$$

$$\boxed{(1) \Rightarrow E[z_0^* z_1] - a E[|z_0|^2] - b E[z_0^*] = 0}$$

$$(2) \Rightarrow E[z_1] - a E[z_0] - \underbrace{E[b]}_{} = 0$$

expectation of a constant is equal to constant

(2) $\times\; E[z_0^*]$

$$= \boxed{E[z_1]\,E[z_0^*] - a\,E[z_0]\,E[z_0^*] - b\,E[z_0^*] = 0} \quad -(3)$$

(1) $-$ (3)

$$\Rightarrow \boxed{E[z_0^* z_1] - E[z_1]\,E[z_0^*] - a\left(E\left[|z_0|^2\right] - a\,E[z_0]\,E[z_0^*]\right) = 0} \quad -(4)$$

In this equation we have the possibility of solving for 'a'

Before doing so lets highlight a property of Covariance which is useful.

$$\gamma_{\alpha\beta} = E\left[(\alpha - m_\alpha)^*(\beta - m_\beta)\right]$$

$$\Rightarrow E\left[(\alpha^* - m_\alpha^*)(\beta - m_\beta)\right]$$

$$\Rightarrow E\left[\alpha^*\beta - \alpha^* m_\beta - m_\alpha^*\beta + m_\alpha^* m_\beta\right]$$

$$\Rightarrow E[\alpha^*\beta] - m_\beta E[\alpha^*] - m_\alpha^* E[\beta] + m_\alpha^* m_\beta$$

$$\Rightarrow E[\alpha^*\beta] - m_\beta m_\alpha^* - \cancel{m_\alpha^* m_\beta} + \cancel{m_\alpha^* m_\beta}$$

$$\Rightarrow E[\alpha^*\beta] - m_\beta m_\alpha^*$$

$\therefore$ Covariance $= \boxed{\begin{aligned} &E[(\alpha - m_\alpha)^*(\beta - m_\beta)] \\ &= E[\alpha^*\beta] - m_\alpha^* m_\beta \end{aligned}}$

$$\rightarrow \underbrace{E[z_0^* z_1] - E[z_1]E[z_0^*]}_{\gamma_{z_0 z_1}} - a\,\underbrace{\left(E[|z_0|^2] - E[z_0]E[z_0^*]\right)}_{a\,\sigma^2_{z_0}} = 0$$

$$\therefore \gamma_{z_0 z_1} = \sigma^2_{z_0} \quad \text{if } z_0 = z_1$$

$$\therefore \gamma_{z_0 z_1} - a\,\sigma^2_{z_0} = 0$$

$$\Rightarrow \boxed{a = \frac{\gamma_{z_0 z_1}}{\sigma^2_{z_0}}}$$

optimal

$$\boxed{b = m_{z_1} - m_{z_0} \frac{r_{z_0 z_1}}{\sigma^2_{z_0}}}$$

predictors