Lib 10/31

Continious Random Variables

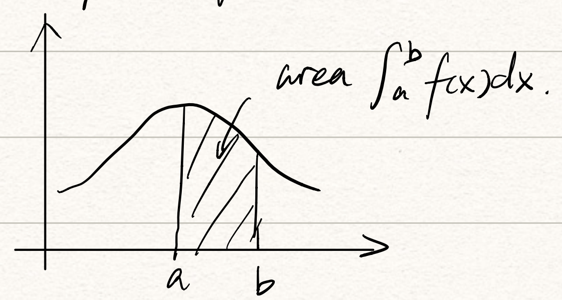
Def: a random variable X is continious if its set of possible values is an entire interval of numbers.

e.g. Ctn r.v. arises from
- physical phenomena - waiting time
  - lifetime of a component.
- limits of disclosure model.

Def. Probability Density Function

Let X be a continious r.v. then a probability density function (pdf) of X is a function $f(x)$ s.t. for any 2 # a, b, with $a \le b$.

$P(a \le X \le b) = \int_a^b f(x)\,dx.$



· For $f(x)$ to be a pdf, $f(x)$ must satisfy:

1) (non-negative) $f(x) \geq 0$, for all x.
2) (total 1) $\int_{-\infty}^{\infty} f(x)\,dx = 1$ total area is 1.

· Comparison with discrete pmf: $p(x)$.
- · $p(x) \geq 0$ for all x.
- · $\sum_i p(x_i) = 1$.

* $f(x)$ probability density, not probability.

$f(x)dx \to$ the probability of $x \in$ a small interval (dx) around x.

a small (infinitesimal) interval near x.

$(x, x+dx)$ or $(x-dx, x)$.

L17 11/5

Exam 2 : 3.2~3.6

3.3 up to square root row.

3.5 up to " sum of independent Poisson distribution are poisson

3.6 hypergeometric distribution p241 ~ 243.

§7.4, §4.1, 4.2 (TBD).

Lec. 10~18.

· $f(x)$ pdf.

$p(a \le x \le b) = \int_a^b f(x)dx$.

E.g. Let x = the "time headway for two randomly selected consecutive cars on a freeway during a period of heavy flow (in secs). Suppose the pdf of x is given.

$$f(x) = \begin{cases} 0.15e^{-0.15(x-0.5)}, & x \ge 0.5 \\ 0, & x < 0.5 \end{cases}$$

Find the probability that the "time headway" is between 1 sec and 2 secs (inclusive).

$$P(1 \le x \le 2) = \int_1^2 f(x) = \int_1^2 0.15e^{-0.15(x-0.5)}dx = -e^{-0.15(x-0.5)}\Big|_1^2$$

$$= e^{-0.15 \cdot 1.5} + e^{-0.15 \cdot 0.5}$$

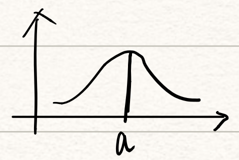
$$\approx 0.1292.$$

Probability of a point

If X is a continuous random variable, then the probability of X will assume any fixed value is 0.

$P(X=a)=0$, for any $a \in \mathbb{R}$.

Proof: $P(X=a) = P(a \le X \le a) = \int_a^a f(x)\,dx = 0.$

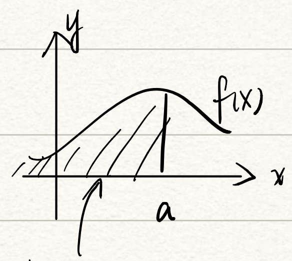


← area of a verticle line $=0$.

Thus $P(a \le x \le b) = P(a<X<b) = P(a \le X < b) = P(a < X \le b)$.

Def: the cultimate distribution function (cdf) $F(x)$ for a continuous r.v. X is defined for every number x by $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$, $-\infty \le x \le \infty$

For each value of x, $F(x)$ = area under the curve ($f(x)$) to the left of x.



The probability of observing X a value less than or equal to x.

$F(a) = P(X \le a)$

Note that: $P(X \le a) = P(X<a) = \int_{-\infty}^{a} f(x)\,dx$.

E.g. 2 Find the cdf of the pdf in E.g. 1.

$$f(x) = \begin{cases} 0.15 e^{-0.15(x-0.5)} & x \ge 0.5 \\ 0 & x < 0.5. \end{cases}$$

$$F(x) = P(X \le x) = \int_{0.5}^{x} 0.15 e^{-0.15(x-0.5)}\,dx + \int_{-\infty}^{0.5} 0\,dx.$$

$= -e^{-0.15(x-0.5)} \Big|_{0.5}^{x} = -e^{-0.15(x-0.5)} + 1, \quad x \geq 0.5.$

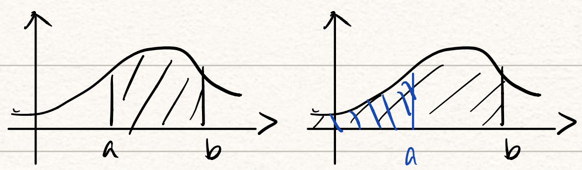
Thus $F(x) = \begin{cases} 1 - e^{-0.15(x-0.5)} & x \geq 0.5 \\ 0 & x < 0.5 \end{cases}$

Properties of cdf.

① compute probabilities using $F(x)$

$P(a \leq x \leq b) = F(b) - F(a)$

Proof: $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$



② obtain pdf $f(x)$ from cdf $F(x)$.

If X is a continuous r.v. with cdf $F(x)$ differentiable at every point x, then the pdf $f(x) = F'(x)$

$P(x \leq X \leq x + dx) = F(x+dx) - F(x) = f(x)dx.$

infinitesimal interval of length dx.

Def: Percentiles

Let $0 \leq p \leq 1$, the $(100p)$th percentile of the distribution of a continuous r.v. X, denoted by $\eta(p)$, is defined by.

$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(x)dx = P(X \leq \eta(p))$

The median of a continuous distribution is the 50^{th} percentile, denoted by $\tilde{\mu}$,
$F(\tilde{\mu}) = 0.5$.

E.g. The waiting time for a bus at a stop (in mins) satisfies the following distribution.

$$f(x) = \begin{cases} \dfrac{x-20}{25}, & 20 \le x \le 25 \\[2mm] \dfrac{30-x}{25}, & 25 \le x \le 30. \end{cases}$$

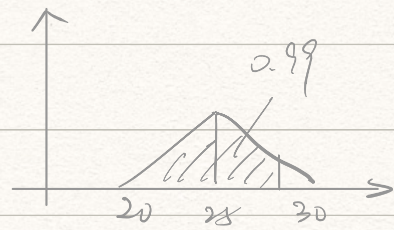
Find the 99^{th} percentile of this distribution.

$F(\eta) = 0.99$

$$\int_{-\infty}^{\eta} f(x)\,dx = \int_{-\infty}^{20} 0\,dx + \int_{20}^{\eta} f(x)\,dx = 0.99$$

$$0.01 = \int_{\eta}^{30} \frac{30-x}{25}\,dx$$

$$= -\frac{(30-x)^2}{50} \Big|_{\eta}^{30}$$



$$0.5 = -0 + (30-\eta)^2$$

$\eta = 30 \pm \sqrt{5}$. $\qquad \eta < 30$

$\eta \approx 29.29$

Def: the expectation (or mean) of a continuous r.v. X with pdf $f(x)$ is $E(X) = \mu_X = \int_{-\infty}^{\infty} x f(x)\,dx$.

[Recall: discrete $E(X) = \sum_{\text{all } x} x\,p(X=x)$].

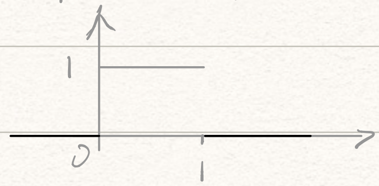
· If X is a continuous r.v. with pdf $f(x)$ and $h(x)$ is a function of x, then $E(h(X)) = \mu_{h(X)} = \int_{-\infty}^{\infty} h(x) f(x)\,dx$.

In particular, if $h(x)=x^2$, then $E(x^2)=\int_{-\infty}^{\infty} x^2 f(x)\,dx$.

Note: if X only takes values from a to b, then
$E(X)=\int_a^b x f(x)\,dx$.

E.g. The probability density function of X is given by

$$f(x)=\begin{cases} 1 & \text{for } 0\le x\le 1 \\ 0 & \text{otherwise.}\end{cases}$$



a) find $E(X)$
b) find $E(e^x)$

Sol. a). $E(X)=\int_0^1 x f(x)\,dx$
$=\int_0^1 x\,dx = \frac{x^2}{2}\Big|_0^1=\frac{1}{2}$

b) $E(e^x)=\int_0^1 e^x f(x)\,dx = e^x\Big|_0^1 = e-1$

E.g. the pdf of the waiting time (in mins) at checkout is given by
$f(x)=\frac{x}{8}$, for $0\le x\le 4$.
a) Find the probability of waiting time less than 3 minutes.
b) Find the expected waiting time.

Sol: a) $P(X<3)=\int_{-\infty}^{3} f(x)\,dx = \int_0^3 f(x)\,dx$
$=\int_0^3 \frac{x}{8}\,dx$
$=\frac{x^2}{16}\Big|_0^3 = \frac{9}{16}$.

b) $E(X)=\int_{-\infty}^{\infty} x f(x)\,dx = \int_0^4 x\cdot\frac{x}{8}\,dx = \frac{x^3}{24}\Big|_0^4 = \frac{8}{3}$.

L18 11/07

E.g. Suppose that if you are $\underline{s}$ minutes early for an appointment, then you incur the cost $\underline{cs}$, and if you are s minutes late, the you incur the cost ks. Suppose that the travel time from where you are presently are to the location of your appointment is a continuous r.v. with pdf f. Determine the time (t) at which you depart if you want to minimize your expected cost.

step 1. Find the expected cost. (a func of time depart & travel time)

X - travel time w/ pdf $f(x)$

t - depart t minutes before the appoinment

cost ~ $h(x,t)$ if s min early, cost $= cs$

s min late, cost $= ks$.

- If travel time (x) less than t, early or late ?

$x < t$ early $(t-x)$ mins, cost $c(t-x)$

$x \geq t$, late $(x-t)$ min, cost $k(x-t)$.

$$h(x,t) = \begin{cases} c(t-x), & x < t \\ k(x-t), & x \geq t. \end{cases}$$

$$E(h(x,t)) = \int_{-\infty}^{\infty} h(x,t) f(x)\, dx.$$

$$L(t) = E(h(x,t)) = \int_0^t c(t-x) f(x)\, dx + \int_t^{\infty} k(x-t) f(x)\, dx.$$

step 2. find t s.t. $\min E(h(x,t))$.

$$\frac{\partial E(h(x,t))}{\partial t} = 0 \Rightarrow \text{solve for } t.$$

$$\frac{\partial}{\partial t} \int_0^t c(t-x) f(x)\, dx = \int_0^t \frac{\partial}{\partial t} [c(t-x) f(x)]\, dx + c\overset{0}{\cancel{(t-x)}} f(x) \Big|_{x=t}$$

$$= \int_0^t c f(x)\, dx.$$

$$\frac{\partial}{\partial t}\int_t^\infty k(x-t)f(x)\,dx = \int_t^\infty \frac{\partial}{\partial t}[k(x-t)f(x)]\,dx - k\overset{0}{\cancel{(x-t)}}f(x)\big|_{x=t}$$
$$= -\int_t^\infty k f(x)\,dx.$$

$$\frac{\partial h(t)}{\partial t} = c\int_0^t f(x)\,dx - k\int_t^\infty f(x)\,dx = 0 \quad \text{Solve for } t.$$

$$\begin{cases} cP(X \le t) - kP(X \ge t) = 0. \\ P(X \le t) + P(X \ge t) = 1 \end{cases}$$

$$P(X \le t) = \frac{k}{c+k}.$$

$F(t) = \frac{k}{c+k} \Rightarrow t$ is the $100\left(\frac{k}{c+k}\right)$ th percentile of the dist of X.

$t = \left(100\frac{k}{c+k}\right)$ th percentile minimize minimize the expected cost.

Recall $\frac{\partial}{\partial t}\int_a^t h(x,t)\,dx = \int_a^t \frac{\partial}{\partial t}h(x,t)\,dx + h(t,t)$

$\frac{\partial}{\partial t}\int_t^b h(x,t)\,dx = \int_t^b \frac{\partial}{\partial t}h(x,t)\,dx - h(t,t)$

One important formula

For non-negative random variable X.

$$E(X) = \int_0^\infty P(X > x)\,dx$$

Recall: X- discrete X-counting r.v. $\{0,1,2,\cdots,n\}$.

$$E(X) = \sum_{j=1}^{n} P(X \ge j)$$

E.g. Suppose $X_1, \cdots, X_n$ iid random variables with the following p.d.f.

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $W = \min(X_1, \cdots, X_n)$ Find the expectation of W.

$E(W) \overset{def}{=}$ w/o pdf

$E(W) = \int_0^1 P(W > x)\,dx.$

$$P(W > x) = P(\min(X_1, \cdots, X_n) > x).$$
$$= P(X_1 > x, X_2 > x, \cdots, X_n > x).$$
$$= P(X_1 > x)\,P(X_2 > x) \cdots P(X_n > x)$$
$$= P(X_1 > x)^n$$
$$= \left(\int_x^1 1\,dt\right)^n$$
$$= (1-x)^n.$$

Thus: $E(W) = \int_0^1 (1-x)^n\,dx = -\frac{(1-x)^{n+1}}{n+1}\Big|_0^1 = \frac{1}{n+1}.$

Variance of continuous r.v. X.

Def. the variance of a ctn random variable X with pdf $f(x)$ and mean μ, is

$$Var(X) = \sigma_X^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$
$$= E[(X-\mu)^2].$$

"mean squared deviation".

Def: SD(X) The standard deviation of a ctn random variable X w/ pdf $f(x)$ is

$$SD(X) = \sigma_X = \sqrt{Var(X)} = \sqrt{E[(X-\mu)^2]}$$

$Var(X) \geq 0 \qquad SD(X) \geq 0.$

An important formula: $Var(X) = EX^2 - (EX)^2$

$$= \int_{-\infty}^{\infty} x^2 f(x)\,dx - \left[\int_{-\infty}^{\infty} x f(x)\,dx\right]^2.$$

$$E(X^2) \geq (EX)^2.$$

E.g. the pdf of waiting time (in mins).

$$f(x) = \frac{x}{8}, \quad 0 \le x \le 4.$$

Find the variance and sd of the waiting time.

$$Var(X) = \int_0^4 x^2 f(x)\,dx - \left(\int_0^4 x f(x)\,dx\right)^2.$$

$$= \int_0^4 x^2 \frac{x}{8}\,dx - \left(\int_0^4 x \cdot \frac{x}{8}\,dx\right)^2$$

$$= \frac{x^4}{32}\Big|_0^4 - \left(\frac{x^3}{24}\Big|_0^4\right)^2$$

$$= \left(\frac{4}{32}\right)^4 - \left(\frac{4^3}{24}\right)^2.$$

$$= \frac{8}{9}.$$

$$SD(X) = \sqrt{\frac{8}{9}} = \frac{2\sqrt{2}}{3}.$$

Property: Transformed random variable.

If $Z = aX + b$, then $E(Z) = aEX + b$.

$$Var(Z) = a^2 Var(X)$$

$$SD(Z) = |a|\, SD(X)$$

(same property holds for discrete cases).

E.g. Suppose X has probability density

$$f(x) = \begin{cases} \frac{1}{(1+x)^2}, & x > 0 \\ 0, & x \le 0. \end{cases}$$

a) Find $P(X > 3)$

b) Let $X_1, \dots, X_4$ are iid rvs w/ same pdf as X.
Find the probability that exactly 2 random variables are greater than 3.

c) Find $E(X)$

Sol. a) $P(X>3) = \int_3^{\infty} f(x)\,dx = \int_3^{\infty} \frac{1}{(1+x)^2}dx$.

$= \lim_{b\to\infty} \int_3^b f(x)\,dx$ $\quad\quad \overset{\text{\textquotesingle\textquotesingle}}{=} -\frac{1}{1+x}\Big|_3^{\infty}$.

$= \lim_{b\to\infty} \left(-\frac{1}{1+x}\Big|_3^b\right)$ $\quad\quad = 0 - (-\frac{1}{4})$

$= \lim_{b\to\infty} (-\frac{1}{1+b}) - (-\frac{1}{4})$ $\quad\quad = \frac{1}{4}$.

$= \frac{1}{4}$.

b) $X_1, X_2, X_3, X_4 \rightarrow$ two of $(X_1, \cdots, X_4) > 3$

$Y \sim Bin(4, \frac{1}{4})$ $\quad \binom{4}{2}$ - # of ways.

$\binom{4}{2} P(\cdot\cdot > 3)^2 \cdot [1 - P(\cdot\cdot > 3)]^2$

$\binom{4}{2} (\frac{1}{4})^2 \cdot (\frac{3}{4})^2 = 0.21$

c) $E(X) = \int_0^{\infty} x f(x)\,dx = \int_0^{\infty} x \frac{1}{(1+x)^2}\,dx$

$= \int_0^{\infty} \frac{x}{(1+x)^2}\,dx$

$= \int_0^{\infty} \left[\frac{x+1}{(1+x)^2} - \frac{1}{(1+x)^2}\right]dx$

$= \int_0^{\infty} \frac{1}{1+x}\,dx - \int_0^{\infty} \frac{1}{(1+x)^2}\,dx$

$= \ln|1+x|\Big|_0^{\infty} - (-\frac{1}{1+x})\Big|_0^{\infty}$

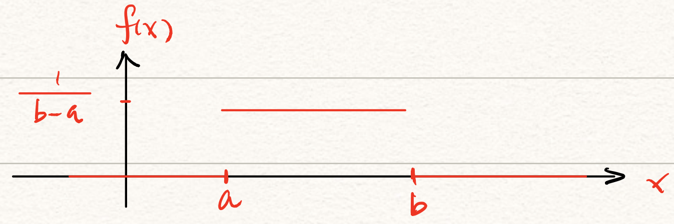
$= \infty - 0 - 0 + (-1) = \infty$.

$E(X) = \infty$, $E(X)$ does not exist!

Def: uniform distribution

A continuous r.v. X is said to have a uniform distribution on the interval (a, b) if the pdf of X is $f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a<x<b \\ 0, & \text{otherwise.} \end{cases}$

Graph.



$X \sim \text{uniform}(a,b)$

If $X \sim \text{uniform}(a,b)$, then

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)\,dx = \frac{x_2 - x_1}{b-a}, \text{ for } a < x_1 < x_2 < b.$$

probability reduces to relative length.

The cdf of $X \sim \text{uniform}(a,b)$ is

$$F(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x \ge b. \end{cases}$$

why? $x \le a$ $F(x) = \int_{-\infty}^{x} 0 \cdot dt = 0.$

$a < x < b$ $F(x) = \int_a^x \frac{1}{b-a}\,dt = \frac{x-a}{b-a}.$

$x \ge b$ $F(x) = \int_a^x f(t)\,dt = \int_a^b f(t)\,dt + \int_b^\infty \underbrace{\cancel{f(t)}}_{0}\,dt = 1.$

Expetation and Variance.

If $X \sim \text{Uniform}(a,b)$, then

- $E(X) = \frac{a+b}{2}$
- $Var(X) = \frac{(b-a)^2}{12}$

11/12 L19

proof: $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_a^b x \cdot \frac{1}{b-a}\,dx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{b+a}{2}$

↑ midpoint of the interval (a,b)

$Var(X) = EX^2 - (EX)^2$

$= \int_a^b x^2 \frac{1}{b-a}\,dx - \left(\frac{a+b}{2}\right)^2$

$$= \frac{x^3}{3(b-a)}\Big|_a^b - \frac{(a+b)^2}{4}$$

$$= \frac{b^3-a^3}{3(b-a)} - \frac{(a+b)^2}{4}$$

$$= \frac{(b-a)(a^2+b^2+ab)}{3(b-a)} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

Alternate proof. ① First find $Z \sim$ uniform $(0,1)$ $\Rightarrow$. $E(Z) =$ _ $Var(X) =$ ____

② Then find $X \sim$ uniform (a,b) (§4.1).

③ transform from x to z

$$z = \frac{x-a}{b-a}$$

E.g. uniform distribution over an area.

Suppose a point is uniformly distributed at random on a circular plate of radius 1. Let R be the distance of the point from the center of the plate.

a) find the probability density of R.

b) find the mean and variance of R.

Sol: pdf $f(r)$ def $\quad P(a<X<b) = \int_a^b f(x)\,dx$.

other way? $\quad \hookrightarrow$ proportion of the value of $R \leq r$.

$$F(r) = P(R \leq r) = \frac{\pi r^2}{\pi} = r^2.$$

$$f(r) = F'(r) = r^{2\prime} = 2r.$$

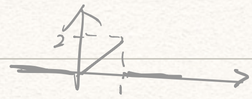
$$f(r) = \begin{cases} 2r & 0 \leq r \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(b) $E(R) = \int_0^1 r f(r)\,dr = \int_0^1 r \cdot 2r\,dr = 2 \cdot \frac{r^3}{3}\Big|_0^1 = \frac{2}{3}$.

$Var(R) = E(R^2) - ER^2 = \int_0^1 r^2 \cdot 2r\,dr - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$.

R ~ uniform over the circular plate.

$f(r) = \begin{cases} 2r & 0 \le r \le 1 \\ 0 & \text{otherwise} \end{cases}$



Def Normal Distribution

A continuous r.v. X is said to have a normal distribution with parameters μ and σ^2 if the pdf of X is $f(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $-\infty < x < \infty$.

This is indeed a pdf { non-negative $f(x) \ge 0$; total one $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Notation $X \sim N(\mu, \sigma^2)$

mean, variance $E(X) = \mu$ $\quad -\infty < \mu < \infty$ → location parameter

$Var(X) = \sigma^2 \ge 0$

$SD(X) = \sigma$ $\quad \sigma \ge 0$ scale parameter.

$=$ iff $P(X=\mu) = 1$

Linear Transformation

If $X \sim N(\mu, \sigma^2)$, then $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$.

Y is also a normal distribution

Standard Normal Distribution

If $\mu = 0$ and $\sigma = 1$, then the normal distribution is called a standard normal distribution, denoted by $Z \sim N(0,1)$, the pdf of r.v. Z.

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad -\infty < z < \infty.$$

The cdf of r.v. Z, is denoted by $\Phi(z)$.

$$\underset{\substack{\downarrow \\ \text{standard normal table.}}}{\Phi(z)} = P(Z \le z) = \int_{-\infty}^{z} f(t)\,dt.$$

Standardization

The transformation from X to Z.

$$Z = \frac{X-\mu}{\sigma}, \quad \text{where } X \sim N(\mu, \sigma^2)$$

is called standardardization

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

Find probabilities

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

100^{th} percentile of X: η

$\eta = \eta_Z \cdot \sigma + \mu$

where η_Z is 100p-th percentile for $Z \sim N(0,1)$

step 1. find η_Z

step 2. Transform: $\eta = \eta_Z \sigma + \mu$.

E.g. the median of $Z \sim N(0,1)$

step 1: η_Z s.t. $\Phi(\eta_Z) = 0.5$ $\quad \eta_Z = 0$.

step 2: $\eta_X = 0 \cdot \sigma + \mu = \mu$ $\quad$ median = mean for $N(\mu, \sigma^2)$ [b/c symmetry]

E.g. Suppose that a binary message, either 0 or 1, must be transmitted by wire from A to B. However, the data sent over the wire are subject to a channel noise disturbance. To reduce the possibility of error, the value 2 is sent when the message 1, and value -2 is sent when ... 0.

· If $X \pm 2$ is the value sent at location A, then R is the value received at location B, is given by $R = X + Z$, where $Z \sim N(0,1)$, is the distribution

· The receiver decodes the message according to the following

$$\begin{cases} \text{if } R \geq .5, \text{ then } 1 \text{ is concluded.} \\ R < 0.5, \text{ then } 0 \text{ is } \ldots \end{cases}$$

Determine the probability of making an error

Sol:

data	A	B	Decodes
1	2	$R = 2 + Z$	$R \geq 0.5 \to 1$
0	-2	$R = -2 + Z$	$< 0.5 \to 0$.

$$\begin{aligned} P(\text{error}) &= P(X=2, R<0.5) + P(X=-2, R \geq 0.5) \\ &= P(2 + Z < 0.5) + P(-2 + Z \geq 0.5) \\ &= P(Z < -1.5) + P(Z \geq 2.5) \\ &= \Phi(-1.5) + 1 - \Phi(2.5) \\ &= 2 - \Phi(1.5) - \Phi(2.5) \\ &\approx 0.73 = 7.3\% . \end{aligned}$$