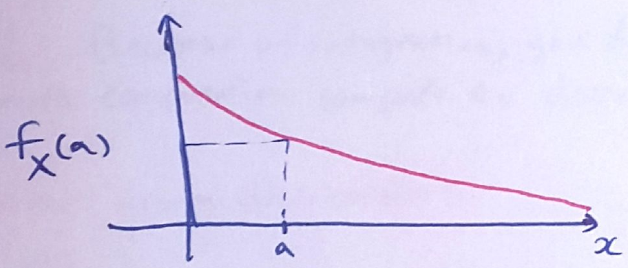
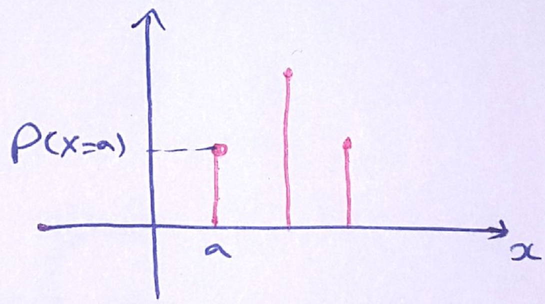
Chapter 5: Continous Random Variables

It is a random variable which can take Uncountable numbers as its value.

X r.v: $X \in \{1, 2, \dots\}$ Infinite elements but countable $\Rightarrow$ Discrete Random Variable

$X \in [0-1]$ " " and uncountable $\Rightarrow$ Continous Random Variable

Continous vs Discrete:

Probability Distribution Function "PDF"	Probability Mass Function "PMF"
$f_X(x)$ $X \in$ Uncountable set (B) e.g. $[2, 4)$, $[0, 4]$	$P(X=x)$ $X \in$ countable set (A) e.g. $\{1,2,3\}$ or $\{1, 4, 16, \dots\}$
$\int_{x \in B} f_X(x)\,dx = 1$	$\sum_{x \in A} P(X=x) = 1$
	
Probability that $X=a \neq f_X(a)$	Probability that $x=a = P(x=a)$
☆ probability can not be defined for a point in continous case	
☆ We only can compute for a range of values.	

Cumulative Distribution Function

"CDF"

$X \in [a,b]$

$$F(x) = \int_a^x f_X(t)\,dt$$

Cumulative Mass Function

"CMF"

$$F(x) = \sum_{k=\min(X)}^{k=x} P(X=k)$$

$$E(x) = \int_a^b x\, f_X(x)\,dx$$

$$E(x) = \sum_{x=\min(X)}^{x=\max(X)} x\, P(X=x)$$

If X is non-negative we also can use:

$$E(x) = \int_a^b [1 - \underbrace{F(x)}_{\text{CDF}}]\,dx$$

(Sometimes is useful in advanced courses)

$$E(x) = \sum_{\min(X)}^{\max(X)} [1 - F(x)]$$

$$E[g(x)] = \int_a^b g(x)\, f_X(x)\,dx$$

$$E[g(x)] = \sum g(x)\, P(X=x)$$

Good: You dont need to find out which distribution should be used. Because it ~~should~~ will be mentioned in question.

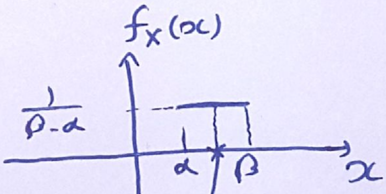
Bad: Because of integration, you deal with more computation compare to discrete case

Good: $\sum$ instead of $\int$

Bad: Sometimes is not easy to find out which distribution you must use.

Some known distributions:

1. Uniform: $X \sim U([\alpha, \beta])$



$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & \alpha \leq x \leq \beta \\ 0 & \text{elsewhere} \end{cases}$$

$$F(c) = \begin{cases} 0 & c \leq \alpha \\ \frac{c-\alpha}{\beta-\alpha} & \alpha < c < \beta \\ 1 & c \geq \beta \end{cases}$$

$$E[x] = \underbrace{\frac{\alpha+\beta}{2}}_{\text{Mid-point}} \qquad Var(x) = \frac{(\beta-\alpha)^2}{12}$$

2. Normal (Gaussian) distribution:

$X \sim N(\mu, \sigma^2)$, where $\mu \to E(X)$ and $\sigma^2 \to Var(X)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}}$$

But if we want to find its CDF by using this PDF, we need complicated integrals to compute

Instead of doing this:

1. Define a random variable called Standard Normal Distribution which is $N(\overset{\mu}{0}, \overset{\sigma^2}{1})$

2. we find a table for its CDF in different points. Call it $\Phi(y)$

3. For a given normal r.v, we transfer that into standard form and use that table

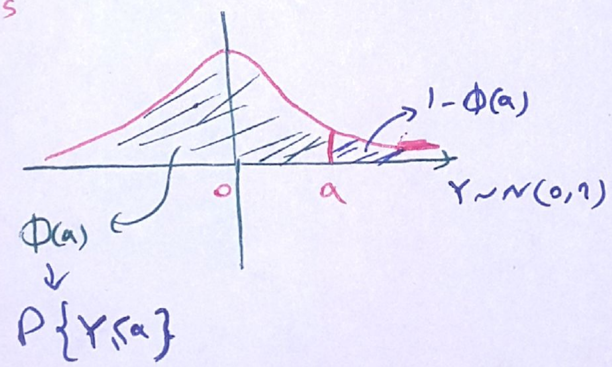
For transfering: $\underbrace{Y}_{\text{Standard}} = \frac{X-\mu}{\sigma}$

$$E[Y] = E\left[\frac{X}{\sigma} - \frac{\mu}{\sigma}\right] = \frac{E[X]}{\sigma} - \frac{E[\mu]}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0 \checkmark$$

$$Var[Y] = Var\left[\frac{X-\mu}{\sigma}\right] = \left(\frac{1}{\sigma^2}\right) Var(X) = \frac{\sigma^2}{\sigma^2} = 1 \checkmark$$

So Y is standard Normal r.v.

For Standard Normal r.v's



3. Exponential random variable

$X \sim Exp(\lambda)$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad E\{x\} = \frac{1}{\lambda} \qquad Var(x) = \frac{1}{\lambda^2}$$

$$F(a) = P(X \leq a) = 1 - e^{-\lambda a} \quad a \geq 0$$

★★ The one and the only memoryless continous random variable

If Lifetime $\sim exp(\lambda)$ already

$P(\text{Lifetime} > 8 \text{ years} \mid \text{Lifetime} > 5) = P(\text{Life time} > 8-5)$ $\quad P\{X > t+s \mid X > t\} = P(X > s)$

Problem 1:

Let X be a continous r.v. having CDF $F_X(x)$. Define the r.v. $Y = F_X(x)$. Show that Y is uniformly distributed over $(0,1)$

Problem 2:

Suppose that we are writing a code in a programming language which only is able to generate uniform random variables. How we can generate an exponential random variable in that?

Problem 3:

Suppose that $X_1, X_2, \dots, X_n$ are independent exponential random variables $(X_i \sim \exp(\mu_i))$ find the PDF of: $\min\{X_1, X_2, \dots, X_n\}$

Problem 4:

Suppose that the cumulative distribution function of the random variable X is given by: $F(x) = 1 - e^{-x^2}$ $x > 0$

Evaluate: a) $P\{X > 2\}$ b) $P\{1 < X < 3\}$ c) the hazard rate function of F

d) $E\{X\}$

problem 5:

<u>EX. 1</u>: You arrive at a bus stop at 10 O'clock, knowing that the bus will arrive at some time uniformly distributed between 10–10:30.

a) what is the prob. that you will have to wait longer than <u>10</u> minutes?

b) If, at 10:15, the bus has not yet arrived, what is the prob. that you will have to wait at least an additional <u>10</u> minutes?

Problem 6:

<u>EX. 3</u>: Let X be a r.v. with the following pdf : $f(x) = \begin{cases} c(1-x^2) & ; -1<x<1 \\ 0 & ; \text{o.w.} \end{cases}$

a) what is the value of $\underline{c}$?

b) Find the prob. that $\underline{X}$ is a positive number?

b) Find the prob. that X is a negative number?

c) Find the variance of $(2\underline{X}+5)$

d) Find the CDF of $\underline{X}$?

Problem 7:

(a) A fire station is to be located along a road of length $A, A < \infty$. If fires occur at points uniformly chosen on $(0, A)$, where should the station be located so as to minimize the expected distance from the fire? That is, choose a so as to

$$\text{minimize } E[|X - a|]$$

when X is uniformly distributed over $(0, A)$.

(b) Now suppose that the road is of infinite length—stretching from point 0 outward to ∞. If the distance of a fire from point 0 is exponentially distributed with rate λ, where should the fire station now be located? That is, we want to minimize $E[|X - a|]$, where X is now exponential with rate λ.

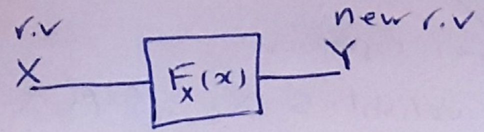
Problem 8:

The lifetimes of interactive computer chips produced by a certain semiconductor manufacturer are normally distributed with parameters $\mu = 1.4 \times 10^6$ hours and $\sigma = 3 \times 10^5$ hours. What is the approximate probability that a batch of 100 chips will contain at least 20 whose lifetimes are less than 1.8×10^6?

Problem 1:

$Y = g(x)$

$g(x) = F_X(x)$

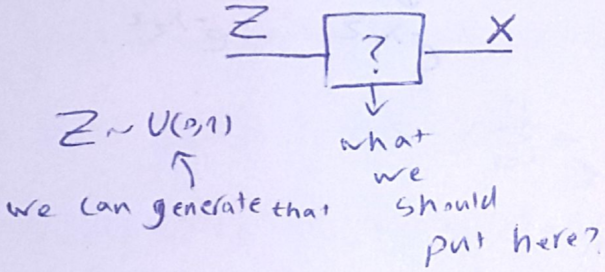


$f_Y(y) = ?$ First we try to find CDF of Y and after that find PDF by taking derivative (In most cases we find PDF from CDF)

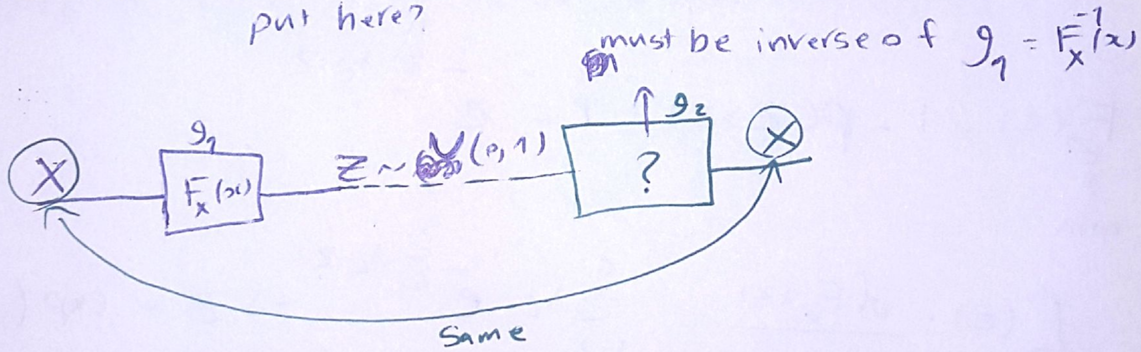
we want to find: $F_Y(y) = P(Y \leq y) \overset{Y = F_X(x)}{=} P(F_X(x) \leq y) = P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y$

$f_Y(y) = \frac{d F_Y(y)}{dy} = \frac{d(y)}{dy} = 1 \Rightarrow Y \sim U(0,1)$ because we know that Y can be from 0 to 1.

Problem 2:



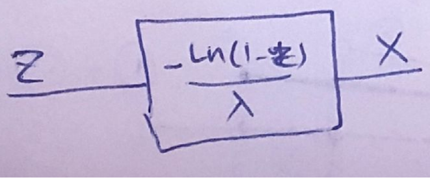
we know that:



So for generating exponential random variable: 1. First we generate uniform random variable for 0 to 1 2. Pass those numbers throug a function which is inverse of the CDF of our objective random variable. In this case: CDF $= F_X(x) = 1 - e^{-\lambda x}$ (exponential)

Finding inverse:

$1 - e^{-\lambda x} = y \Rightarrow 1 - y = e^{-\lambda x} \Rightarrow \ln(1-y) = -\lambda x \Rightarrow x = \frac{-\ln(1-y)}{\lambda} = F_X^{-1}(x)$



☆ Note: CDF is always inversable ☆

Problem 3:

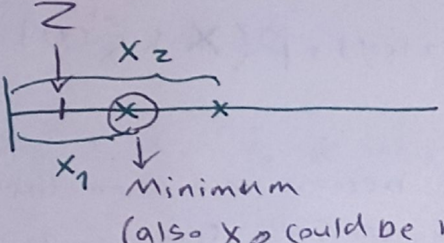
$$Z = \min\{X_1, X_2, \dots, X_n\}$$

Independent exponential random variables $X_i \sim \exp(\lambda_i)$

we want to find: $f_Z(z)$

we start with CDF: $P(Z \leq z) = \underbrace{1 - P(Z > z)}$

For one reason it is better to use this



$$P(\min\{X_1, X_2, \dots, X_n\} > z)$$

$$= P(X_1 > z \cap X_2 > z \cap \dots X_n > z)$$

$$\overset{\text{Independency}}{=} \underbrace{P(X_1 > z)}_{e^{-\lambda_1 z}} \underbrace{P(X_2 > z)}_{e^{-\lambda_2 z}} \cdots \underbrace{P(X_n > z)}_{e^{-\lambda_n z}}$$

$$\Rightarrow P(\min > z) = \prod_{i=1}^{n} e^{-\lambda_i z} = e^{-\sum_{i=1}^{n} \lambda_i z}$$

$$F_Z(z) = 1 - P(Z > z) = 1 - e^{-\sum_{i=1}^{n} \lambda_i z}$$

$\uparrow$ min

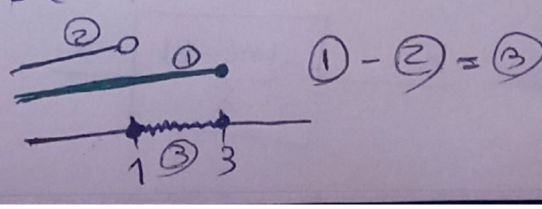
$$f_Z(z) = \frac{dF_Z(z)}{dz} = \sum_{i=1}^{n} \lambda_i \, e^{-\sum_{i=1}^{n} \lambda_i z} \Rightarrow Z \sim \exp\left(\sum_{i=1}^{n} \lambda_i\right)$$

Problem 4:

$$F_X(x) = \begin{cases} 1 - e^{-x^2} & 0 \leq x < \infty \\ 0 & x < 0 \end{cases}$$

a) $P\{X > 2\} = 1 - P\{X \leq 2\} = 1 - F_X(2) = 1 - (1 - e^{-2^2}) = e^{-4}$

b) $P\{1 \leq X \leq 3\} = P\{X \leq 3\} - P\{X < 1\}$



c)

Hazard rate function $h(x)$ or $\lambda(x)$ $= \dfrac{f_X(x)}{\bar{F}_X(x)} = \dfrac{f_X(x)}{1-F_X(x)} = \dfrac{2x e^{-x^2}}{1-(1-e^{-x^2})} = 2x$

$$f_X(x) = \frac{dF_X(x)}{dx} = 2x e^{-x^2}$$

d)

$$E\{X\} = \int_0^\infty x f_X(x)\,dx = \int_0^\infty x(2x e^{-x^2})\,dx = 2\int_0^\infty x^2 e^{-x^2}\,dx = I$$

Integration by parts:

$2x e^{-x^2}dx = dv \Rightarrow v = -e^{-x^2}$

$x = u \Rightarrow du = 1\,dx$

$\Rightarrow I = uv\Big|_0^\infty - \int_0^\infty v\,du$

$I = (-x e^{-x^2})\Big|_0^\infty - \int_0^\infty (-e^{-x^2})\,dx$

$$I = (0-0) + \int_0^\infty e^{-x^2}\,dx$$

To find this we know: $\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{\infty} e^{-\frac{s^2}{2}}\,ds = 1$

$S \sim N(0,1)$

Changing ~~[illegible]~~ Variable: $\dfrac{s}{\sqrt{2}} = x$

$\Downarrow$

$ds = \sqrt{2}\,dx$

$\Rightarrow \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{\infty} e^{-x^2}\sqrt{2}\,dx = 1$

$\Downarrow$

$\dfrac{2}{\sqrt{\pi}}\displaystyle\int_0^\infty e^{-x^2}\,dx = 1.$

$\Downarrow$

$\displaystyle\int_0^\infty e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}$

$$\Rightarrow \boxed{I = \frac{\sqrt{\pi}}{2}}$$

Scanned by CamScanner

Problem 5:

You arrive at a bus stop at 10 O'clock, knowing that the bus will arrive at some time uniformly distributed between 10–10:30.

a) what is the prob. that you will have to wait longer than 10 minutes?

Let's define $X = \#$ of minutes past 10 o'clock, when the bus arrives.

Therefore, X has a uniform distn, i.e., $X \sim U[0,30]$, $f_X(x) = \begin{cases} \frac{1}{30} & ; 0<x<30 \\ 0 & ; \text{O.W.} \end{cases}$

$P\{X>10\} = \int_{10}^{30} \frac{1}{30}\,dx = \frac{20}{30} = \frac{2}{3}$, Note: if $X \sim U[a,b] \rightarrow f_X(x) = \begin{cases} \frac{1}{b-a} & ; a\leq x\leq b \\ 0 & ; \text{O.W.} \end{cases}$

b) If, at 10:15, the bus has not yet arrived, what is the prob. that you will have to wait at least an additional 10 minutes?

$$P(X > 15+10 \mid X>15) = \frac{P\{X>25, X>15\}}{P\{X>15\}} = \frac{P\{X>25\}}{P\{X>15\}} = \frac{\int_{25}^{30} \frac{1}{30}\,dx}{\int_{15}^{30} \frac{1}{30}\,dx} = \frac{5/30}{15/30} = \frac{1}{3}$$

Problem 6:

Let X be a normal random variable with mean 12 and variance 4. Find the value of $\underline{C}$ such that $P\{X>C\} = 0.1$?

$X \sim N(12,4) \rightarrow Y = \frac{X-12}{2} \sim N(0,1)$

$P\{X > C\} = P\{\underbrace{\frac{X-12}{2}}_{Y} > \frac{C-12}{2}\} = 1 - P\{Y \leq \frac{C-12}{2}\} = 1 - \Phi(\frac{C-12}{2}) = 0.1$

$\rightarrow \Phi(\frac{C-12}{2}) = 0.9 \longrightarrow$ from the table (textbook) : $\Phi(1.28) \simeq 0.9$

$\longrightarrow \frac{C-12}{2} = 1.28 \longrightarrow C = 14.56$

EX.3 : Let X be a r.v. with the following pdf : $f(x) = \begin{cases} c(1-x^2) & ; -1<x<1 \\ 0 & ; \text{O.W.} \end{cases}$

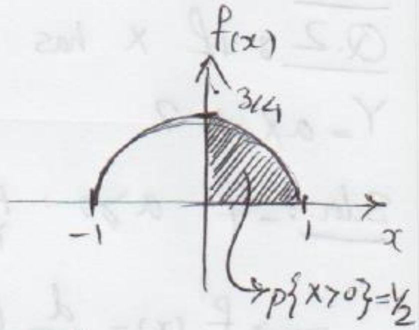
a) what is the value of c?

$\int_{-1}^{1} f_X(x)dx = 1 \longrightarrow \int_{-1}^{+1} c(1-x^2)dx = c\left(x - \frac{x^3}{3}\right)\Big|_{-1}^{+1} = c\left[(1-1/3) - (-1+1/3)\right] = c \cdot 4/3 = 1$

$\longrightarrow c = 3/4 \longrightarrow f_X(x) = \begin{cases} 3/4(1-x^2) & ; -1<x<1 \\ 0 & ; \text{O.W.} \end{cases}$

b) Find the prob. that $\underline{X}$ is a positive number?

$$p\{X>0\} = \int_0^1 \tfrac{3}{4}(1-x^2)\,dx = \tfrac{3}{4}\left(x-\tfrac{x^3}{3}\right)\Big|_0^1 = \tfrac{3}{4}\left(1-\tfrac{1}{3}\right) = \tfrac{1}{2}$$



b) Find the prob. that X is a negative number?

* The curve of pdf is symmetric: $p\{X<0\} = \tfrac{1}{2} = 1 - p\{X>0\}$

* Can also compute $p\{X<0\} = \int_{-1}^{0} f_X(x)\,dx = \tfrac{1}{2}$

c) Find the variance of $(2\underline{X}+5)$? $Var(2X+5) = 4\,Var(X)$

$$Var(X) = E(X^2) - (E(X))^2 \;;\; E(X) = \int_{-1}^{+1} x\cdot f_X(x)\,dx = \int_{-1}^{+1} \tfrac{3}{4}(x-x^3)\,dx = \tfrac{3}{4}\left(\tfrac{x^2}{2}-\tfrac{x^4}{4}\right)\Big|_{-1}^{1} = 0$$

$$E(x^2) = \int_{-1}^{+1} x^2 \cdot f_X(x)\,dx = \int_{-1}^{+1} \tfrac{3}{4}(x^2-x^4)\,dx = \tfrac{3}{4}\left(\tfrac{x^3}{3}-\tfrac{x^5}{5}\right)\Big|_{-1}^{+1} = \tfrac{1}{5}$$

$$\longrightarrow Var(2X+5) = 4\cdot\left[\tfrac{1}{5} - 0\right] = \tfrac{4}{5}$$

d) Find the CDF of $\underline{X}$?

$$F_X(x) = p(X \leq x) = \int_{-1}^{x} f_X(x)\,dx = \int_{-1}^{x} \tfrac{3}{4}(1-x^2)\,dx = \tfrac{3}{4}\left(x - \tfrac{x^3}{3} + \tfrac{2}{3}\right), \; -1<x<1$$

Problem 7:

Part (a): If x (the location of the fire) is uniformly distributed in $[0, A)$ then we would like to select a (the location of the fire station) such that

$$F(a) \equiv E[\,|X-a|\,]\,,$$

is a minimum. We will compute this by breaking the integral involved in the definition of the expectation into regions where $x-a$ is negative and positive. We find that

$$\begin{aligned}
E[\,|X-a|\,] &= \int_0^A |x-a|\frac{1}{A}dx \\
&= -\frac{1}{A}\int_0^a (x-a)dx + \frac{1}{A}\int_a^A (x-a)dx \\
&= -\frac{1}{A}\left.\frac{(x-a)^2}{2}\right|_0^a + \frac{1}{A}\left.\frac{(x-a)^2}{2}\right|_a^A \\
&= -\frac{1}{A}\left(0-\frac{a^2}{2}\right) + \frac{1}{A}\left(\frac{(A-a)^2}{2}-0\right) \\
&= \frac{a^2}{2A} + \frac{(A-a)^2}{2A}\,.
\end{aligned}$$

To find the a that minimizes this we compute $F'(a)$ and set this equal to zero. Taking the derivative and setting this equal to zero we find that

$$F'(a) = \frac{a}{A} + \frac{2(A-a)(-1)}{2A} = 0\,.$$

Which gives a solution a^* given by $a^* = \frac{A}{2}$. A second derivative of our function F shows that $F''(a) = \frac{2}{A} > 0$ showing that the point $a^* = A/2$ is indeed a minimum.

Part (b): The problem formulation is the same as in part (a) but since the distribution of the location of fires is now an exponential we now want to minimize

$$F(a) \equiv E[\,|X-a|\,] = \int_0^\infty |x-a|\lambda e^{-\lambda x}dx\,.$$

We will compute this by breaking the integral involved in the definition of the expectation into regions where $x-a$ is negative and positive. We find that

$$E[\,|X-a|\,] = \int_0^\infty |x-a|\lambda e^{-\lambda x}dx$$

$$
\begin{aligned}
&= -\int_0^a (x-a)\lambda e^{-\lambda x}dx + \int_a^\infty (x-a)\lambda e^{-\lambda x}dx \\
&= -\lambda\left(\frac{(x-a)}{-\lambda}e^{-\lambda x}\Big|_0^a + \frac{1}{\lambda}\int_0^a e^{-\lambda x}dx\right) \\
&+ \lambda\left(\frac{(x-a)}{-\lambda}e^{-\lambda x}\Big|_a^\infty + \frac{1}{\lambda}\int_a^\infty e^{-\lambda x}dx\right) \\
&= -\lambda\left(\frac{-a}{\lambda} - \frac{1}{\lambda^2}e^{-\lambda x}\Big|_0^a\right) + \lambda\left(0 - \frac{1}{\lambda^2}e^{-\lambda x}\Big|_a^\infty\right) \\
&= a + \frac{1}{\lambda}(e^{-\lambda a} - 1) - \frac{1}{\lambda}(-e^{-\lambda a}) \\
&= a + \frac{1+2e^{-\lambda a}}{\lambda}.
\end{aligned}
$$

To find the a that minimizes this we compute $F'(a)$ and set this equal to zero. Taking the derivative we find that

$$F'(a) = 1 - 2e^{-\lambda a} = 0\,.$$

Problem 8:

If each chips lifetime is denoted by the random variable X (assumed Gaussian with the given mean and variance), then each chip will have a lifetime less than $1.8\,10^6$ hours with probability given by

$$
\begin{aligned}
P\{X < 1.8\,10^6\} &= P\left\{\frac{X - 1.4\,10^6}{3\,10^5} < \frac{(1.8-1.4)\,10^6}{3\,10^5}\right\} \\
&= P\left\{Z < \frac{4}{3}\right\} = \Phi(4/3) \approx 0.9088\,.
\end{aligned}
$$

With this probability, the number N, in a batch of 100 that will have a lifetime less than $1.8\,10^6$ is a binomial random variable with parameters $(n,p) = (100, 0.9088)$. Therefore, the probability that a batch will contain at least 20 is given by

$$P\{N \geq 20\} = \sum_{n=20}^{100} \binom{100}{n} (0.908)^n (1-0.908)^{100-n}\,.$$

Rather than evaluate this exactly we can approximate this binomial random variable N with a Gaussian random variable with a mean given by $\mu = np = 100(0.908) = 90.87$, and a variance given by $\sigma^2 = np(1-p) = 8.28$ (equivalently $\sigma = 2.87$). Then the probability that a given batch of 100 has at least 20 that have lifetime less than $1.8\,10^6$ hours is given by

$$
\begin{aligned}
P\{N \geq 20\} &= P\{N \geq 19.5\} \\
&= P\left\{\frac{N - 90.87}{2.87} \geq \frac{19.5 - 90.87}{2.87}\right\} \\
&\approx P\{Z \geq -24.9\} \\
&= 1 - P\{Z \leq -24.9\} \\
&= 1 - \Phi(-24.9) \approx 1\,.
\end{aligned}
$$

Where in the first line above we have used the continuity correction required when we approximate a discrete density by a continuous one, and in the third line above we use our Gaussian approximation to the binomial distribution.