Common mistakes in homeworks

1. Probabilities are <u>numbers</u>. <u>Allowed</u> operations: $+, -, \times, \%$

Right: $P(A \cup B)$ ✓ Wrong: $P(A) \cup P(B)$ ✗

Events are <u>NOT</u> numbers. Allowed operations: $\cup, \cap, {}^c$

Right: $P(A) + P(B)$ ✓ Wrong: $P(A + B - C)$ ✗

2. <u>Only</u> for independent events

$$P(A \cap B) = P(A) \cdot P(B) \leftarrow$$ if you use this, check that events are independent; state it.

For <u>all</u> events:

$$P(A \cap B) = P(A|B) \cdot P(B)$$

3. When proving something, you cannot assume what you're trying to prove and arrive at what's given

Say A, B are indep, prove A and B^c are indep

<u>Wrong</u>: $P(A \cap B^c) = P(A) \cdot P(B^c) = P(A)(1 - P(B)) = P(A) - P(A) \cdot P(B)$

In these notes

1. Normal random variables
   - ex 3 from lecture 4 slides
   - H/w 2, question 4 a) & 4 b)

2. ~~[illegible]~~ Random variable transformations
   - Tutorial 2, question 2.3

3. Poisson / exponential variables
   - Tutorial 2, question 2.1

Normal distribution

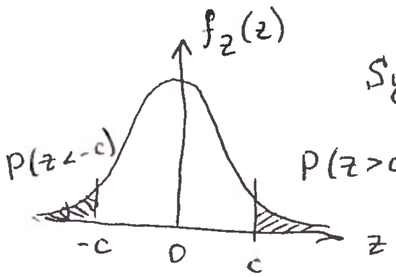
1. Standard Normal distribution

$$Z \sim N(0,1)$$

$$\mathbb{E}(Z) = 0 \quad , \quad Var(Z) = 1$$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \Rightarrow F_Z(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = \Phi(z)$$

→ cannot be solved in close form



Symmetric, bell - shaped, unimodal

$$P(Z>c) = P(Z<-c) \Rightarrow 1-\Phi(c) = \Phi(-c)$$

Values of Φ can be obtained via pnorm in R

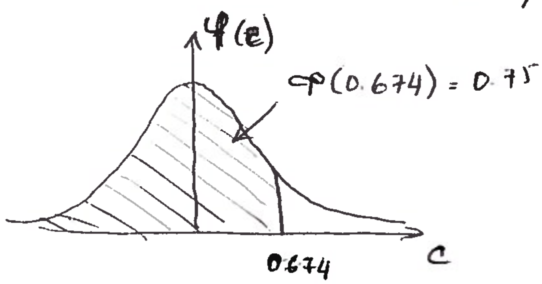
$\Phi(2) = P(Z \le 2) = 0.9772$ $\boxed{\text{pnorm}(2)}$

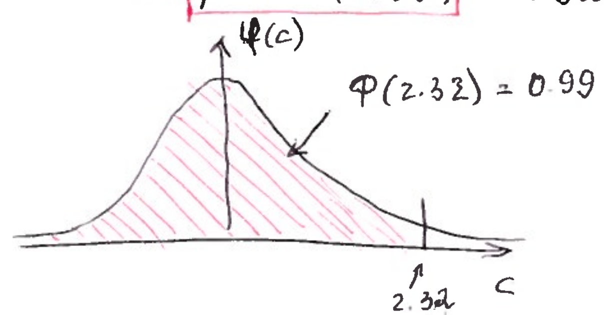
$\Phi(0) = P(Z \le 0) = 0.5$ $\boxed{\text{pnorm}(0)}$

Quantiles of standard Normal distrib. can be obtained in R

$\Phi(c) = 0.75 \Rightarrow P(Z \le c) = 0.75$ $\quad c = \boxed{\text{qnorm}(0.75)} = 0.674$

$\Phi(c) = 0.99 \Rightarrow P(Z \le c) = 0.99$ $\quad c = \boxed{\text{qnorm}(0.99)} = 2.32$





2. Non - standard normal

$$X \sim N(\mu, \sigma^2) \quad \text{if} \quad X = \mu + \sigma Z$$

$$F_X(c) = P(X \le c) = P\left(\frac{X-\mu}{\sigma} \le \frac{c-\mu}{\sigma}\right) = P\left(Z \le \frac{c-\mu}{\sigma}\right) = \Phi\left(\frac{c-\mu}{\sigma}\right)$$

Properties: if $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ and X_1, X_2 <u>independent</u>

- $X_1 \pm X_2 \sim N(\mu_1 \pm \mu_2, \sigma_1^2 \underline{\underline{+}} \sigma_2^2)$ ← for both $X_1 + X_2$ and $X_1 - X_2$

- $aX_1 \pm bX_2 \sim N(a\mu_1 \pm b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$

Important: $Var(aX_1) = a^2 Var(X_1) = a^2\sigma_1^2$

$Var(X_1 \pm X_2) = Var(X_1) + Var(X_2) = \sigma_1^2 + \sigma_2^2$

Normal distribution example

Recall: $Z = \frac{X-\mu}{\sigma}$

$\mu = \mathbb{E}(X)$, $\sigma^2 = Var(X)$

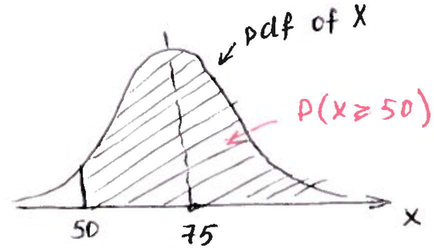
1. Made-up question.

The distribution of test scores is normal with mean 75 and standard deviation 10. a) What is the probability that a student passes the test?

Let X = test score of the student

$X \sim N(\mu = 75, \sigma^2 = 10^2)$

$\sim$ "follows distribution"



$P(\text{student passes}) = P(X \geq 50)$

$= P\left(\frac{X-\mu}{\sigma} \geq \frac{50-\mu}{\sigma}\right)$

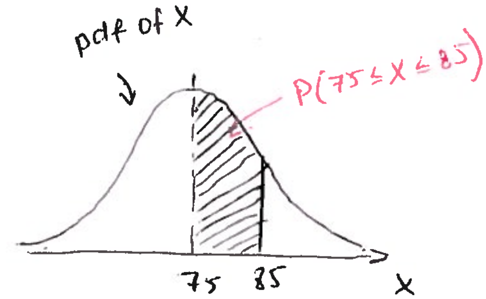
$(\mu = 75, \sigma = 10)$ $= P\left(Z \geq \frac{50-75}{10}\right)$

$= 1 - P(Z < -2.5)$

in R

$= 1 - \Phi(-2.5) = 1 - \boxed{\text{pnorm}(-2.5)} = 0.994$

b) What is the probability that the student scores between 75 and 85 on the test



$P(75 \leq X \leq 85) = P\left(\frac{75-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{85-\mu}{\sigma}\right)$

$= P(0 \leq Z \leq 1)$

$= \Phi(1) - \Phi(0)$

from R $= \boxed{\text{pnorm}(1)} - \text{pnorm}(0)$

= 50% because $Z \sim N(0,1)$ is symmetric around 0

$= 0.3413$

2. Problem 3 from lecture 4

A machine fills "10-pound" bags of concrete mix. The actual weight of the bag is a normal R.V. with SD (stand dev). $\sigma = 0.1$ lbs. The mean is set by the operator

Let X = weight of a bag

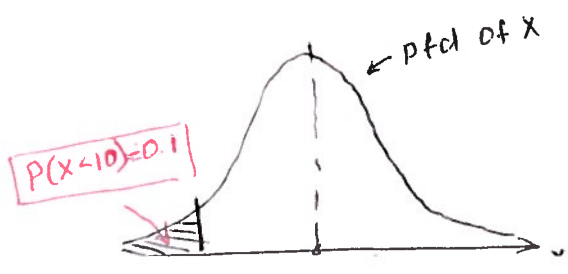
$X \sim N(\mu, 0.1^2)$ variance = $(SD)^2$

a) What μ do we need if at most 10% of bags should be underfilled

"underfilled" = less than 10 pounds

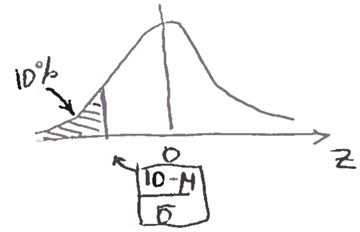
$P(X < 10) \leq 0.1$

$P\left(\frac{X-\mu}{\sigma} < \frac{10-\mu}{\sigma}\right) \leq 0.1$



$P\left(z < \frac{10-\mu}{\sigma}\right) \leq 0.1$

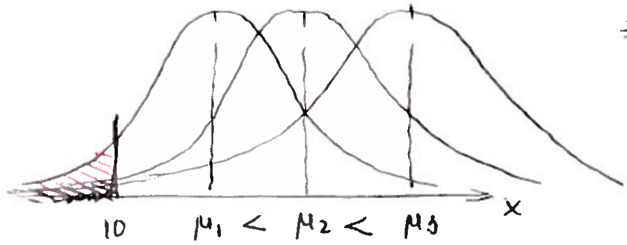
use $\boxed{\text{qnorm}(0.1)} = -1.2816$



$(\sigma = 0.1)$ $\quad \frac{10-\mu}{0.1} = -1.2816$

$\hookrightarrow$ solve for μ: $\mu = 10.128$

Bigger μ provides smaller $P(x<10)$, see plot



$\Rightarrow$ the Bigger the μ the smaller the $P(X<10)$

$\Rightarrow$ For $P(X<10) \leq 0.1$

$\mu \geq 10.128$

b) Now define $\sigma = 0.1\mu$. Now $X \sim N(\mu, 0.1^2\mu^2)$

$P(x<10) \leq 0.1$ the question is the same

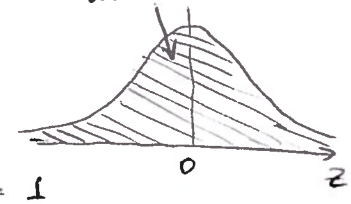
$P\Big(\underbrace{\frac{x-\mu}{\sigma}}_{=z} < \frac{10-\mu}{\sigma}\Big) \leq 0.1$, $\quad P\left(z < \frac{10-\mu}{0.1\mu}\right) \leq 0.1$

Same, use $\boxed{\text{qnorm}(0.1)} = -1.2816$

$\frac{10-\mu}{0.1\mu} = -1.2816 \quad \underset{\text{solve for } \mu}{\Rightarrow} \quad \mu \geq 11.470$

Assignment 2, problem 4

Proving that this whole area is 1



a) $Z \sim N(0,1)$ - standard normal

pdf: $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, $\quad -\infty < z < +\infty$

Show that $I = \int_{-\infty}^{+\infty} \varphi(z)\,dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-z^2/2}\,dz = 1$

/This asks us to prove that Normal density integrates to 1, as it should because it is a density/

Notice:

$$I^2 = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,dx \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\,dy =$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(\frac{1}{\sqrt{2\pi}}\right)^2 e^{-\frac{x^2}{2}-\frac{y^2}{2}}\,dx\,dy = \frac{1}{2\pi} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{(x^2+y^2)}{2}}\,dx\,dy \qquad (*)$$

Change of variables into polar coordinates:

Let $x = r\cos v$, $y = r\sin v$, $r > 0$, $v \in [-\pi, \pi]$

Compute the Jacobian:

$$|J| = \begin{vmatrix} \frac{dx}{dr} & \frac{dx}{dv} \\ \frac{dy}{dr} & \frac{dy}{dv} \end{vmatrix} = \begin{vmatrix} \cos v & -r\sin v \\ \sin v & r\cos v \end{vmatrix}$$

Recall: $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$ (determinant of 2x2 matrix)

$$= r\cos v \cdot \cos v - (-r\sin v \cdot \sin v)$$
$$= r(\underbrace{\cos^2 v + \sin^2 v}_{=1})$$
$$= r$$

Back to the integral $(*)$

$$(*) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \int_{0}^{+\infty} e^{-\frac{(r\cos v)^2 + (r\sin v)^2}{2}} |J| \, dr \, d\theta$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \int_{0}^{+\infty} e^{-r^2/2} \cdot r \, dr \cdot d\theta = \frac{1}{2\pi} \underbrace{\int_{-\pi}^{\pi} d\theta}_{=2\pi} \cdot \int_{0}^{+\infty} e^{-r^2/2} r\,dr$$

Substitution
$t = -r^2/2$
$dt = d(-r^2/2)$
$dt = -\frac{\cancel{2}r}{\cancel{2}} dr = -r\,dr$

$$= \int_{0}^{+\infty} e^{-r^2/2} r\,dr$$

$$= -\int_{0}^{+\infty} e^{t} dt$$

plug $t = -r^2/2$

$$= -e^{t}\Big|_{0}^{+\infty} = -e^{-r^2/2}\Big|_{0}^{+\infty} = -\lim_{r\to\infty} e^{-r^2/2} - (-e^{0}) = 0 + 1 = 1$$

b) $\mathbb{E}(Z)$ - similar substitution to the one above

By def: $\mathbb{E}(Z^2) = \int_{-\infty}^{+\infty} z^2 \varphi(z)\, dz$, where $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$

Hint: $\varphi'(z) = -z\varphi(z) \Rightarrow z^2\varphi(z) = -z\varphi'(z)$ plug into the integral.

Substitution
$d\varphi(z) = \varphi'(z)\,dz$

$$\mathbb{E}(Z^2) = -\int_{-\infty}^{+\infty} z\,\varphi'(z)\,dz$$

$$= -\int_{-\infty}^{+\infty} z\,d\varphi(z) = -z\varphi(z)\Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \varphi(z)\,dz = 0 + 1 = 1$$

integral of normal density is 1 by (a)

plug in $\varphi(z)$ and calculate limits

Random variable transformations

Type of questions: when given RV (random variable) X with some distribution and $y = g(x)$ – R.V. Y is some function of X

Asked to find CDF or PDF or mean or other characteristic of Y.

Algorithm to solve these questions

1. Find the CDF of X (if not already given)

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt, \quad f_X(t) \text{ - pdf of } X$$

2. Write the expression for CDF of Y and sub $g(x)$ instead of Y

$$F_Y(y) = P(Y \le y) = P(g(X) \le y)$$

3. Solve the inequality inside the probability for X and plug the result into $F_X(x)$

$$P(g(X) \le y) = P(X \le g^{-1}(y)) = F_X(g^{-1}(y))$$

(4.) Determine the domain of $F_Y(y)$, i.e the range of Y

Example Tutorial 2, problem 2.3 X - R.V. $F_X(x) = 1 - \frac{1}{x^2}$, $x > 1$

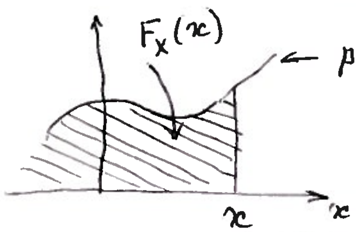
a) find pdf of $Y = X^2$

How to read notation: $F_X(x) = P(X \le x)$ definition of CDF

CDF of X (R.V.), evaluated at x is the probability that X (R.V) takes values less than or equal to x

CDF = cumulative distribution function

pdf = probability density function



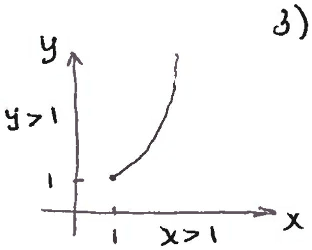
Using algorithm above for $Y = X^2$

1) CDF of X given $F_X(x) = 1 - \frac{1}{x^2}$, $x > 1$

2) $F_Y(y) = P(Y \le y) = P(X^2 \le y)$ since $X > 0$

3) $= P(X \le \sqrt{y})$

$$= 1 - \frac{1}{(\sqrt{y})^2} = 1 - \frac{1}{y}, \quad y > 1$$



Domain: $x > 1$

$y = x^2 > 1$

pdf: $f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}\left(1 - \frac{1}{y}\right) = \begin{cases} \frac{1}{y^2}, & y > 1 \\ 0, & \text{otherwise} \end{cases}$

B) find pdf of $V = \frac{1}{X}$

using same algo as above:

1) CDF of X given $F_X(x) = 1 - \frac{1}{x^2}$, $x > 1$

2) $F_V(v) = P(V \le v) = P\left(\frac{1}{X} \le v\right) \longrightarrow$

$\frac{1}{x} \le v \quad | \cdot x \quad x, v > 0$

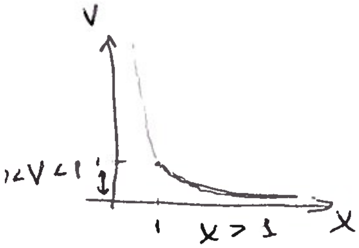
$1 \le vx \quad | \div v$

$\frac{1}{v} \le x$

3) $= P\left(X \ge \frac{1}{v}\right)$

$= 1 - P\left(X < \frac{1}{v}\right)$

$= 1 - \left(1 - \frac{1}{(1/v)^2}\right) = v^2 \quad 0 < v < 1$



Domain: $X > 1$

$0 < V = \frac{1}{X} < 1$

V cannot go below 0

In full form $F_V(v) = \boxed{P(V \le v)} = \begin{cases} 0, & v \le 0 \\ v^2, & 0 < v < 1 \\ 1, & v \ge 1 \end{cases}$

V is always less then or equal to 1

$f_V(v) = \frac{d}{dv} F_V(v) = \begin{cases} 2v, & 0 < v < 1 \\ 0, & \text{otherwise} \end{cases}$

c) find pdf of $W = \ln X$

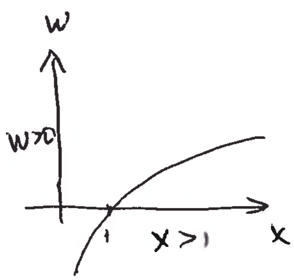
Same algorithm:

1) CDF of X: $F_X(x) = 1 - \frac{1}{x^2}$, $x > 1$

2) $F_W(w) = P(W \le w) = P(\ln X \le w)$

3) $= P(X \le e^w)$

$= 1 - \frac{1}{(e^w)^2} = 1 - e^{-2w}$, $w > 0$



Domain: $X > 1$

$W = \ln X > 0$

$f_W(w) = \frac{d}{dw} F_W(w) = 2e^{-2w}$, $w > 0$

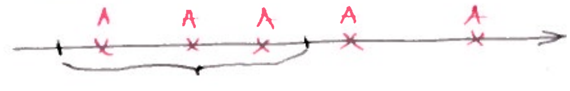
$W \sim \text{Exponential}(\lambda = 2)$

Poisson / exponential random variables

Problem 2.1, Tutorial 2

A happens with rate $\lambda = 3$ per hour.

Define X = # of events during the next hour
$X \sim \text{Pois}(\lambda)$ /has Poisson distribution/



X is discrete
$$\boxed{P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}}$$ $x = 0, 1, 2, \dots$
← p.m.f (prob. mass function of poisson R.V)

Poisson R.V.s measure the # of events during a fixed stretch of time t (or space)

a) What is the P 5 events A happen during next 2 hours

Let X_1 = # of events during next 2 hours
$X_1 \sim \text{Pois}(2\lambda) = \text{Pois}(6)$

$$P(X_1 = 5) = \frac{e^{-6}6^5}{5!} = 0.1606$$

b) What is the P that we have to wait more than 30 mins for the next occurrence of A?

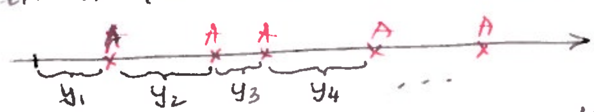
Approach 1: Let X_2 = # of events during next 30 mins ($\frac{1}{2}$ hour)
$X_2 \sim \text{pois}(\frac{1}{2}\lambda) = \text{pois}(1.5)$

{Next A occurs in over 30 mins} = {0 events A in the next 30 min}

$$P(X_2 = 0) = \frac{e^{-1.5}1.5^0}{0!} = \boxed{e^{-1.5}}$$

Approach 2: with exponential random variables

Define y = time until the next event A (or time between events)
$y \sim \text{Exp}(\lambda)$ /has Exponential dist/ ← same λ as Poisson



* y measures time in the same units as λ (i.e. if λ = rate per hour, then y is in hours, if rate per day, then y is in days, etc)

pdf: $f_y(y) = \lambda e^{-\lambda y}, \ y > 0$, y is continuous
CDF: $F_y(y) = P(y \le y) = 1 - e^{-\lambda y}, \ y > 0$

$y \sim \text{Exp}(3)$ - time in hours till next event

$$P(y > \tfrac{1}{2}) = 1 - P(y \le \tfrac{1}{2}) = 1 - F_y(\tfrac{1}{2}) = 1 - (1 - e^{-3 \cdot 1/2}) = \boxed{e^{-1.5}}$$

c) Compute expected time until the next event, $\mathbb{E}(y) =$:

Def Expected value for a continuous R.V. y is

$$\mathbb{E}(y) = \int_{-\infty}^{+\infty} y\, f_y(y)\, dy \quad , \quad f_y(y) \text{ is the pdf of } y$$

$y \sim Exp(\lambda)$

$$\mathbb{E}(y) = \int_{-\infty}^{+\infty} y \cdot \overbrace{\cancel{\lambda e^{-\lambda y}}}^{f_y(y)} dy = \underbrace{\int_{-\infty}^{0} 0\, dy}_{} + \int_{0}^{+\infty} y \cdot \lambda e^{-\lambda y} dy$$

exponential density only for $y > 0 \Rightarrow$ it's 0 elsewhere

Integration by parts:

$$\int_a^b u\,dv = uv\Big|_a^b - \int_a^b v\,du$$

$$= \lambda \int_0^{+\infty} y\, e^{-\lambda y} dy$$

$$e^{-\lambda y} dy = -\frac{1}{\lambda} d e^{-\lambda y}$$

$$-\frac{1}{\lambda} d e^{-\lambda y} = -\frac{1}{\lambda} \cdot e^{-\lambda y} \cdot (-\lambda)\, dy = e^{-\lambda y} dy$$

$$= -\lambda \int_0^{+\infty} y \cdot \frac{1}{\lambda} d e^{-\lambda y}$$

$$= -\frac{\lambda}{\lambda} \int_0^{+\infty} y\, d e^{-\lambda y} = -y e^{-\lambda y}\Big|_0^{+\infty} + \int_0^{+\infty} e^{-\lambda y} dy = \cancel{-\lim_{y \to +\infty} y e^{-\lambda y}} \to 0 + 0 +$$

$$+ \int_0^{+\infty} e^{-\lambda y} dy = -\frac{1}{\lambda} e^{-\lambda y}\Big|_0^{+\infty} = \cancel{-\lim_{y \to +\infty} \frac{1}{\lambda} e^{-\lambda y}} \to 0 + \frac{1}{\lambda} e^{-\lambda \cdot 0} = \frac{1}{\lambda}$$

Or compute Moment Generating function $\mathbb{E}(e^{ty})$

$$M_y(t) = \mathbb{E}(e^{ty}) = \int_0^{+\infty} e^{ty} \lambda e^{-\lambda y} dy = \lambda \int_0^{+\infty} e^{-(\lambda - t) y} dy =$$

$$= -\frac{1}{\lambda - t} \cdot \lambda \cdot e^{-(\lambda - t) y}\Big|_0^{+\infty} =$$

$$= \underbrace{-\lim_{y \to +\infty} \frac{\lambda}{\lambda - t} e^{-(\lambda - t) y}}_{*} + \frac{\lambda}{\lambda - t} e^{-(\lambda - t) \cdot 0}$$

* This limit exists only if $(\lambda - t) > 0$ (and then it's 0).
if $(\lambda - t) < 0 \Rightarrow \lim = \infty \Rightarrow$ MGF does not exist

$$= 0 + \frac{\lambda}{\lambda - t} \text{ only for } \lambda - t > 0, \text{ i.e. } t < \lambda = \boxed{M_y(t) = \frac{\lambda}{\lambda - t}, \; t < \lambda}$$

Can find $\mathbb{E}(y) = \frac{d}{dt} M_y(t)\Big|_{t=0}$

and $\mathbb{E}(y^2) = \frac{d^2}{dt^2} M_y(t)\Big|_{t=0}$